Design Recommendations for Intelligent Tutoring Systems

Volume 2
Instructional Management

Instructional Strategies

Authoring & Expert Modeling

Intelligent Tutoring Systems

Domain Modeling

Learners

Team Tutoring

Learning Effect Evaluations

Learner Modeling

Edited by:
Robert Sottilare
Arthur Graesser
Xiangen Hu
Benjamin Goldberg

A Book in the Adaptive Tutoring Series

Design Recommendations for Intelligent Tutoring Systems

Volume 2
Instructional Management

Edited by:
Robert A. Sottilare, Arthur C. Graesser,
Xiangen Hu, and Benjamin S. Goldberg

A Book in the Adaptive Tutoring Series

Printed in the United States of America
First Printing, June 2014
First Printing (errata addressed), July 2014

*U.S. Army Research Laboratory
Human Research & Engineering Directorate
SFC Paul Ray Smith Simulation & Training Technology Center
Orlando, Florida*

Dedicated to current and future scientists and developers of adaptive learning technologies

CONTENTS

Section IV: Instruction and Scaffolding (Graesser) 259

PREFACE

Robert A. Sottilare[1], Arthur C. Graesser[2], Xiangen Hu[2], and Benjamin S. Goldberg[1]

U.S. Army Research Laboratory - Human Research and Engineering Directorate[1]

University of Memphis Institute for Intelligent Systems[2]

This book is the second in a planned series of books that examine key topics (e.g., learner modeling, instructional strategies, authoring, domain modeling, learning effect, and team tutoring) in intelligent tutoring system (ITS) design through the lens of the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg, and Holden, 2012), a modular, service-oriented architecture created to develop standards for authoring, managing instruction, and analyzing the effect of ITS technologies.

This preface introduces tutoring functions, provides instructional best practices, and examines the motivation for standards for the design, authoring, instruction, and analysis functions within ITSs. Next, we introduce GIFT design principles, and finally, we discuss how readers might use this book as a design tool. We begin by examining the major components of ITSs.

Components and Functions of Intelligent Tutoring Systems

It is generally accepted that an ITS has four major components (Elson-Cook, 1993; Nkambou, Mizoguchi & Bourdeau, 2010; Graesser, Conley & Olney, 2012; Psotka & Mutter, 2008; Sleeman & Brown, 1982; VanLehn, 2006; Woolf, 2009): The domain model, the student model, the tutoring model, and the user-interface model. GIFT similarly adopts this four-part distinction, but with slightly different corresponding labels (domain module, learner module, pedagogical module, and tutor-user interface) and the addition of the sensor module, which can be viewed as an expansion of the user interface.

(1) The domain model contains the set of skills, knowledge, and strategies of the topic being tutored. It normally contains the ideal expert knowledge and also the bugs, mal-rules, and misconceptions that students periodically exhibit.

(2) The learner model consists of the cognitive, affective, motivational, and other psychological states that evolve during the course of learning. It is often viewed as an overlay (subset) of the domain model, which changes over the course of tutoring. For example, "knowledge tracing" tracks the learner's progress from problem to problem and builds a profile of strengths and weaknesses relative to the domain model (Anderson, Corbett, Koedinger & Pelletier, 1995). An ITS may also consider psychological states outside of the domain model that need to be considered as parameters to guide tutoring.

(3) The tutor model (also known as the pedagogical model or the instructional model) takes the domain and learner models as input and selects tutoring strategies, steps, and actions on what the tutor should do next in the exchange. In mixed-initiative systems, the learners may also take actions, ask questions, or request help (Aleven, McClaren, Roll & Koedinger, 2006; Rus & Graesser, 2009), but the ITS always needs to be ready to decide "what to do next" at any point and this is determined by a tutoring model that captures the researchers' pedagogical theories.

(4) The user interface interprets the learner's contributions through various input media (speech, typing, clicking) and produces output in different media (text, diagrams, animations, agents). In addition to the conventional human-computer interface features, some recent systems have incorporated natural language interaction (Graesser et al., 2012; Johnson & Valente, 2008), speech recognition (D'Mello, Graesser & King, 2010; Litman, 2013), and the sensing of learner emotions (Baker, D'Mello, Rodrigo & Graesser, 2010; D'Mello & Graesser, 2010; Goldberg, Sottilare, Brawner, Holden, 2011).

The designers of the tutor model need to decide what best practices of instruction and human tutoring are represented in the model along with methods to select optimal ITS strategies (plans) and tactics (actions) based on the learner's states, traits and data, and the instructional context.

Principles of Learning and Instructional Techniques, Strategies & Tactics

Instructional techniques, strategies, and tactics play a central role in the design of the Generalized Intelligent Framework for Tutoring (GIFT). Instructional techniques represent instructional best practices and principles from the literature many of which have yet to be implemented within GIFT at the writing of this volume. Examples of instructional techniques include, but are not limited to error-sensitive feedback, mastery learning, adaptive spacing and repetition, and fading worked examples. Others are represented in the next section of this preface. It is anticipated that techniques within GIFT will be implemented as software-based agents where the agent will monitor learner progress and instructional context to determine if best practices (agent policies) have been adhered to or violated. Over time the agent will learn to enforce agent policies in a manner that optimizes learning and performance.

As noted many of the best instructional practices (techniques) have yet to be implemented in GIFT, but instructional strategies and tactics have been implemented. Instructional strategies (plans for action by the tutor) are selected based on changes to the learner's state (cognitive, affective, physical). If a sufficient change in any learner's state occurs, this trigger's GIFT to select a generic strategy (e.g., provide feedback). The instructional context along with the instructional strategy then triggers the specific selection of an instructional tactic (an action to be taken by the tutor). If the strategy is "provide feedback", then the tactic might be to "provide feedback on the error committed during the presentation of instructional concept 'B' in the chat window during the next turn." Tactics detail what is to be done, why, when, and how. Additional details on strategies and tactics is discussed in the prologue (Nye, Sottilare, Ragusa & Hoffman) of this volume.

An adaptive, intelligent learning environment needs to launch the right instructional strategies at the right time in a mechanism that attempts to be sensitive to the learner model; maximize learning and motivation; and minimize training time and costs. Instructional management was the theme of the second advisory board meeting of the collaboration between (1) the Human Research and Engineering Directorate (HRED) of the U.S. Army Research Laboratory (ARL) and (2) the Advanced Distributed Learning Center for Intelligent Tutoring Systems Research & Development (ADL CITSRD) in the Institute for Intelligent Systems (IIS) at the University of Memphis. The purpose of this volume is to provide a succinct illustration of some instructional strategies and associated principles of learning in order to orient participants at the board meeting.

Instructional strategies have been advocated by researchers and practitioners in many different fields, such as education, educational psychology, cognitive and learning sciences, military training, computer based training, artificial intelligence in education, computer supported collaborative learning, educational data mining — the list goes on. These fields have different missions, so the shared knowledge among members of different fields is unspectacular. The landscape of instructional strategies in one field would not necessarily overlap with the other fields. However, a common ground has been emerging from dozens of reports prepared by interdisciplinary research panels funded by the government and research organizations, particularly during the last decade. The following are some examples:

- A Roadmap to Educational Technology (2010, National Science Foundation, http://www.cra.org/ccc/docs/groe/GROE Roadmap for Education Technology Final Report.pdf)

- The Army Learning Concept for 2015 (2011, United States Army, http://www-tradoc.army.mil/tpubs/pams/tp525-8-2.pdf)

- Committee on Science Learning: Computer Games, Simulations, and Education (2011, National Academy of Science, http://www.nap.edu/catalog.php?record_id=13078)

- Assessing 21st Century Skills (2011, National Academy of Sciences, http://www.nap.edu/catalog.php?record_id=13215#toc)

- Improving Adult Literacy Instruction (2012, National Academy of Sciences, http://www.nap.edu/catalog.php?record_id=13242)

- Organizing Instruction and Study to Improve Student Learning (2007, Institute of Education Sciences of the United States Department of Education, http://ies.ed.gov/ncee/wwc/pdf/practice_guides/20072004.pdf)

- Lifelong Learning at Work and at Home (2007, American Psychological Association and Association for Psychological Sciences - see Inaugural editorial for Journal of Educational Psychology, http://www.apa.org/pubs/journals/features/edu-101-2-259.pdf),

These reports emphasize instructional strategies that are supported by empirical tests with scientific methodologies. Therefore, the strategies are grounded in science and evidence-based rather than a folklore of educational practitioners. Nevertheless, all of these reports also emphasize practical applications of these strategies. Some reports go to great lengths describing how human teachers can apply particular strategies in teaching practice. Most of them describe computer applications that have implemented and tested the strategies. These reports are, therefore, relevant to GIFT.

Two of these reports illustrate some recommended instructional strategies. *Organizing Instruction and Study to Improve Student Learning* was to serve as a practice guide for teachers. The goal was to focus on a small number of strategies that were backed by science and that could also be reliably applied with the training that teachers typically receive. In other words, the instructional strategies should not be too complex or subtle. The research group identified the following seven principles:

1. Space learning over time.

2. Interleave worked example solutions with problem solving exercises.

3. Combine graphics with verbal descriptions.

4. Connect and integrate abstract and concrete representations of concepts.

5. Use quizzing to promote learning.

6. Help students allocate study time effectively.

7. Ask deep explanatory questions.

Lifelong Learning at Work and at Home had a larger and more diverse panel of experts, with an eye toward adult learners in addition to K-12. These experts generated 25 principles of learning and instructional best practices.

1. **Contiguity Effects**: Ideas that need to be associated should be presented contiguously in space and time.

2. **Perceptual-motor Grounding:** Concepts benefit from being grounded in perceptual motor experiences, particularly at early stages of learning.

3. **Dual Code and Multimedia Effects**: Materials presented in verbal, visual, and multimedia form richer representations than a single medium.

4. **Testing Effect:** Testing enhances learning, particularly when the tests are aligned with important content.

5. **Spacing Effect:** Spaced schedules of studying and testing produce better long-term retention than a single study session or test.

6. **Exam Expectations:** Students benefit more from repeated testing when they expect a final exam.

7. **Generation Effect:** Learning is enhanced when learners produce answers compared to having them recognize answers.

8. **Organization Effects:** Outlining, integrating, and synthesizing information produces better learning than rereading materials or other more passive strategies.

9. **Coherence Effect:** Materials and multimedia should explicitly link related ideas and minimize distracting irrelevant material.

10. **Stories and Example Cases:** Stories and example cases tend to be remembered better than didactic facts and abstract principles.

11. **Multiple Examples:** An understanding of an abstract concept improves with multiple and varied examples.

12. **Feedback Effects:** Students benefit from feedback on their performance in a learning task, but the timing of the feedback depends on the task.

13. **Negative Suggestion Effects:** Learning wrong information can be reduced when feedback is immediate.

14. **Desirable Difficulties:** Challenges make learning and retrieval effortful and thereby have positive effects on long-term retention.

15. **Manageable Cognitive Load:** The information presented to the learner should not overload working memory.

16. **Segmentation Principle:** A complex lesson should be broken down into manageable subparts.

17. **Explanation Effects:** Students benefit more from constructing deep coherent explanations (mental models) of the material than memorizing shallow isolated facts.

18. **Deep Questions:** Students benefit more from asking and answering deep questions that elicit explanations (e.g., why, why not, how, what-if) than shallow questions (e.g., who, what, when, where).

19. **Cognitive Disequilibrium:** Deep reasoning and learning is stimulated by problems that create cognitive disequilibrium, such as obstacles to goals, contradictions, conflict, and anomalies.

20. **Cognitive Flexibility:** Cognitive flexibility improves with multiple viewpoints that link facts, skills, procedures, and deep conceptual principles.

21. **Goldilocks Principle:** Assignments should not be too hard or too easy, but at the right level of difficulty for the student's level of skill or prior knowledge.

22. **Metacognition:** Students rarely have an accurate knowledge of their cognition so their ability to calibrate their comprehension, learning, and memory should not be trusted and they need to be trained to improve important metacognitive judgments.

23. **Discovery Learning:** Most students have trouble discovering important principles on their own, without careful guidance, scaffolding, or materials with well-crafted affordances.

24. **Self-regulated Learning:** Most students need training on how to self-regulate their learning and other cognitive processes.

25. **Anchored Learning:** Learning is deeper and students are more motivated when the materials and skills are anchored in real world problems that matter to the learner.

These lists provide an initial glimpse of instructional strategies, but in a number of ways fall short of providing sufficient guidance for GIFT. The precise conditions in which each strategy should be applied require further specification. Indeed, each strategy is appropriate for some conditions but not others, e.g., distributed over massed practice is typically desirable, but sometimes massed practice is best. There are contradictions or tradeoffs between some of these strategies, e.g., coherence effect versus cognitive disequilibrium. Another shortcoming is that these strategies emphasize cognitive mechanisms at the expense of not giving adequate attention to motivation, emotions, and social interaction. We live in a world where these non-cognitive factors are just as important as cognitive mechanisms.

Members of the second advisory board were selected because their research fills many of these gaps and provides more sophisticated instructional strategies for GIFT. More specifically, researchers on the board have made major advances in four thematic subcategories: (1) meta-cognition and self-regulated learning, (2) affect, emotions, engagement, and grit, (3) guided instruction and scaffolding, and (4) natural language and discourse. Research in these subcategories is destined to move the horizon of instructional strategies beyond conventional computer-based instruction and onto learning environments with serious games, virtual reality, self-regulation, social interaction, and scaffolding techniques for enhancing both learning and motivation.

Motivations for Intelligent Tutoring System Standards

An emphasis on self-regulated learning has highlighted a requirement for point-of-need training in environments where human tutors are either unavailable or impractical. ITSs have been shown to be as effective as expert human tutors (VanLehn, 2011) in one-to-one tutoring in well-defined domains (e.g., mathematics or physics) and significantly better than traditional classroom training environments. ITSs have demonstrated significant promise, but fifty years of research have been unsuccessful in making ITSs ubiquitous in military training or the tool of choice in our educational system. Why?

The availability and use of ITSs have been constrained by their high development costs, their limited reuse, a lack of standards, and their inadequate adaptability to the needs of learners (Picard, 2006). Their application to military domains is further hampered by the complex and often ill-defined environments in which our military operates today. ITSs are often built as domain-specific, unique, one-of-a-kind, largely domain-dependent solutions focused on a single pedagogical strategy (e.g., model tracing or constraint-based approaches) when complex learning domains may require novel or hybrid approaches. Therefore, a modular ITS framework and standards are needed to enhance reuse, support authoring, optimize instruc-

tional strategies, and lower the cost and skillset needed for users to adopt ITS solutions for training and education. It was out of this need that the idea for GIFT arose.

GIFT has three primary functions: authoring, instructional management, and analysis. First, it is a framework for authoring new ITS components, methods, strategies, and whole tutoring systems. Second, GIFT is an instructional manager that integrates selected tutoring principals and strategies for use in ITSs. Finally, GIFT is an experimental testbed to analyze the effectiveness and impact of ITS components, tools, and methods. GIFT is based on a learner-centric approach with the goal of improving linkages in the updated adaptive tutoring learning effect chain (Figure P-1).
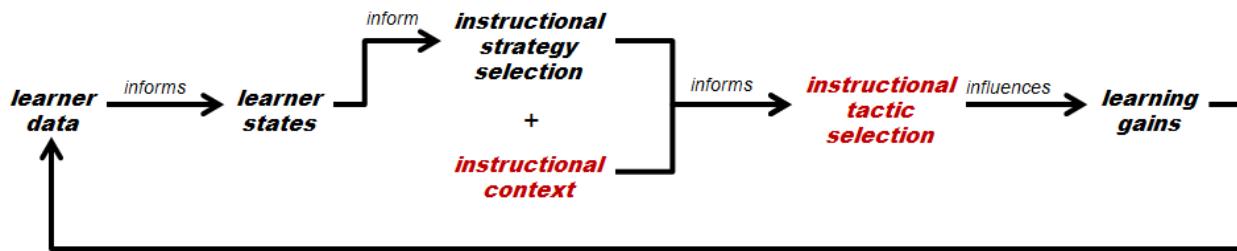
Figure P-1. Adaptive Tutoring Learning Effect Chain

(Sottilare, 2012; Sottilare, Ragusa, Hoffman & Goldberg, 2013)

A deeper understanding of the learner's behaviors, traits, and preferences (learner data) collected through performance, physiological and behavioral sensors, and surveys will allow for more accurate evaluation of the learner's states (e.g., engagement level, confusion, frustration), which will result in a better and more persistent model of the learner. To enhance the adaptability of the ITS, methods are needed to accurately classify learner states (e.g., cognitive, affective, psychomotor, social) and select optimal instructional strategies given the learner's existing states. A more comprehensive learner model will allow the ITS to adapt more appropriately to address the learner's needs by changing the instructional strategy (e.g., content, flow, or feedback). An instructional strategy that is better aligned to the learner's needs is more likely to positively influence their learning gains. It is with the goal of optimized learning gains in mind that the design principles for GIFT were formulated.

GIFT Design Principles

The methodology for developing a modular, computer-based tutoring framework for training and education considered major design goals, anticipated uses, and applications. The design process also looked at enhancing one-to-one (individual) and one-to-many (collective or team) tutoring experiences beyond the state of practice for ITSs today. A significant focus of the GIFT design was on domain-dependent elements in the domain module. This was done to allow large-scale reuse of the remaining GIFT modules across different training domains and thereby reduce the development costs for ITSs.

One design principle adopted in GIFT is that each module should be capable of gathering information from other modules according to the design specification. Designing to this principle resulted in standard message sets and message transmission rules (i.e., request-driven, event-driven, or periodic transmissions). For instance, the pedagogical module is capable of receiving information from the learner module to develop courses of action for future instructional content to be displayed, manage flow and challenge

level, and select appropriate feedback. Changes to the learner's state (e.g., engagement, motivation, or affect) trigger messages to the pedagogical module, which then recommends general courses of action (e.g., ask a question or prompt the learner for more information) to the domain module, which provides a domain-specific intervention (e.g., what is the next step?).

Another design principle adopted within GIFT is the separation of content from the executable code (Patil & Abraham, 2010). Data and data structures are placed within models and libraries, while software processes are programmed into interoperable modules. Efficiency and effectiveness goals (e.g., accelerated learning and enhanced retention) were considered to address the time available for military training and the renewed emphasis on self-regulated learning. An outgrowth of this emphasis on efficiency and effectiveness led Dr. Sottilare to seek external collaboration and guidance. In 2012, U.S. Army Research Laboratory (ARL) with the University of Memphis developed advisory boards of senior tutoring system scientists from academia and government to influence the GIFT design goals moving forward. An advisory board for learner modeling was completed in September 2012, and future boards are planned for instructional strategy design, authoring and expert modeling, learning effect evaluations, and domain modeling.

Design Goals and Anticipated Uses

GIFT may be used as any of the following:

1. An architectural framework with modular, interchangeable elements and defined relationships

2. A set of specifications to guide ITS development

3. A set of exemplars instantiating GIFT to support authoring and ease-of-use

4. A technical platform or testbed for guiding the development of concrete systems

These use cases have been distilled down into the three primary functional areas, or *constructs*: authoring, instructional management, and analysis. Discussed below are the purposes, associated design goals, and anticipated uses for each of the GIFT constructs.

GIFT Authoring Construct

The purpose of the GIFT authoring construct is to provide technology (tools and methods) to make it affordable and easier to build ITSs and ITS components. Toward this end, a set of extensible markup language (XML) configuration tools continues to be developed to allow for data-driven changes to the design and implementation of GIFT-generated ITSs. The design goals for the GIFT authoring construct have been adapted from Murray (1999, 2003) and Sottilare & Gilbert (2011). The GIFT authoring design goals are as follow:

- Decrease the effort (time, cost, and/or other resources) for authoring and analyzing ITSs by automating authoring processes, developing authoring tools and methods, and developing standards to promote reuse.

- Decrease the skill threshold by tailoring tools for specific disciplines (e.g., instructional designers, training developers, and trainers) to author, analyze, and employ ITS technologies.

- Provide tools to aid designers/authors/trainers/researchers in organizing their knowledge.

- Support (structure, recommend, or enforce) good design principles in pedagogy through user interfaces, and other interactions.

- Enable rapid prototyping of ITSs to allow for rapid design/evaluation cycles of prototype capabilities.

- Employ standards to support rapid integration of external training/tutoring environments (e.g., simulators, serious games, slide presentations, transmedia narratives, and other interactive multimedia).

- Develop/exploit common tools and user interfaces to adapt ITS design through data-driven means.

- Promote reuse through domain-independent modules and data structures.

- Leverage open-source solutions to reduce ITS development and sustainment costs.

- Develop interfaces/gateways to widely used commercial and academics tools (e.g., games, sensors, toolkits, virtual humans).

As a user-centric architecture, anticipated uses for GIFT authoring tools are driven largely by the anticipated users, which include learners, domain experts, instructional system designers, training and tutoring system developers, trainers and teachers, and researchers. In addition to user models and graphical user interfaces, GIFT authoring tools include domain-specific knowledge configuration tools, instructional strategy development tools, and a compiler to generate executable ITSs from GIFT components in a variety of formats (e.g., PC, Android, and IPad).

Within GIFT, domain-specific knowledge configuration tools permit authoring of new knowledge elements or reusing existing (stored) knowledge elements. Domain knowledge elements include learning objectives, media, task descriptions, task conditions, standards and measures of success, common misconceptions, feedback library, and a question library, which are informed by instructional system design principles that, in turn inform concept maps for lessons and whole courses. The task descriptions, task conditions, standards and measures of success, and common misconceptions may be informed by an expert or ideal learner model derived through a task analysis of the behaviors of a highly skilled user. ARL is investigating techniques to automate this expert model development process to reduce the time and cost of developing ITSs. In addition to feedback and questions, supplementary tools are anticipated to author explanations, summaries, examples, analogies, hints, and prompts in support of GIFT's instructional management construct.

GIFT Instructional Management Construct

The purpose of the GIFT instructional management construct is to integrate pedagogical best practices in GIFT-generated ITSs. The modularity of GIFT will also allow GIFT users to extract pedagogical models for use in tutoring/training systems that are not GIFT-generated. GIFT users may also integrate pedagogical models, instructional strategies, or instructional tactics from other tutoring systems into GIFT. The design goals for the GIFT instructional management construct are the following:

- Support ITS instruction for individuals and small teams in local and geographically distributed training environments (e.g., mobile training), and in both well-defined and ill-defined learning domains.

- Provide for comprehensive learner models that incorporate learner states, traits, demographics, and historical data (e.g., performance) to inform ITS decisions to adapt training/tutoring.

- Support low-cost, unobtrusive (passive) methods to sense learner behaviors and physiological measures and use these data along with instructional context to inform models to classify (in near real time) the learner's states (e.g., cognitive and affective).

- Support both macro-adaptive strategies (adaptation based on pre-training learner traits) and micro-adaptive instructional strategies and tactics (adaptation based learner states and state changes during training).

- Support the consideration of individual differences where they have empirically been documented to be significant influencers of learning outcomes (e.g., knowledge or skill acquisition, retention, and performance).

- Support adaptation (e.g., pace, flow, and challenge level) of the instruction based the domain and learning class (e.g., cognitive learning, affective learning, psychomotor learning, social learning).

- Model appropriate instructional strategies and tactics of expert human tutors to develop a comprehensive pedagogical model.

To support the development of optimized instructional strategies and tactics, GIFT is heavily grounded in learning theory, tutoring theory, and motivational theory. Learning theory applied in GIFT includes conditions of learning and theory of instruction (Gagne, 1985), component display theory (Merrill, Reiser, Ranney & Trafton, 1992), cognitive learning (Anderson & Krathwohl, 2001), affective learning (Krathwohl, Bloom, and Masia, 1964; Goleman, 1995), psychomotor learning (Simpson, 1972), and social learning (Sottilare, Holden, Brawner, and Goldberg, 2011; Soller, 2001). Aligning with our goal to model expert human tutors, GIFT considers the INSPIRE model of tutoring success (Lepper, Drake, and O'Donnell-Johnson, 1997) and the tutoring process defined by Person, Kreuz, Zwaan, and Graesser (1995) in the development of GIFT instructional strategies and tactics.

INSPIRE is an acronym that highlights the seven critical characteristics of successful tutors: Intelligent, Nurturant, Socratic, Progressive, Indirect, Reflective, and Encouraging. Graesser & Person's (1994) tutoring process includes a tutor-learner interchange where the tutor asks a question, the learner answers the question, the tutor gives feedback on the answer, then the tutor and learner collaboratively improve the quality of (or embellish) the answer. Finally, the tutor evaluates learner's understanding of the answer.

As a learner-centric architecture, anticipated uses for GIFT instructional management capabilities include both automated instruction and blended instruction, where human tutors/teachers/trainers use GIFT to support their curriculum objectives. If its design goals are realized, it is anticipated that GIFT will be widely used beyond military training contexts as GIFT users expand the number and type of learning domains and resulting ITS generated using GIFT.

GIFT Analysis Construct

The purpose of the GIFT analysis construct is to allow ITS researchers to experimentally assess and evaluate ITS technologies (ITS components, tools, and methods). The design goals for the GIFT analysis construct are the following:

- Support the conduct of formative assessments to improve learning

- Support summative evaluations to gauge the effect of technologies on learning

- Support assessment of ITS processes to understand how learning is progressing throughout the tutoring process

- Support evaluation of resulting learning versus stated learning objectives

- Provide diagnostics to identify areas for improvement within ITS processes

- Support the ability to comparatively evaluate ITS technologies against traditional tutoring or classroom teaching methods

- Develop a testbed methodology to support assessments and evaluations (Figure P-2)
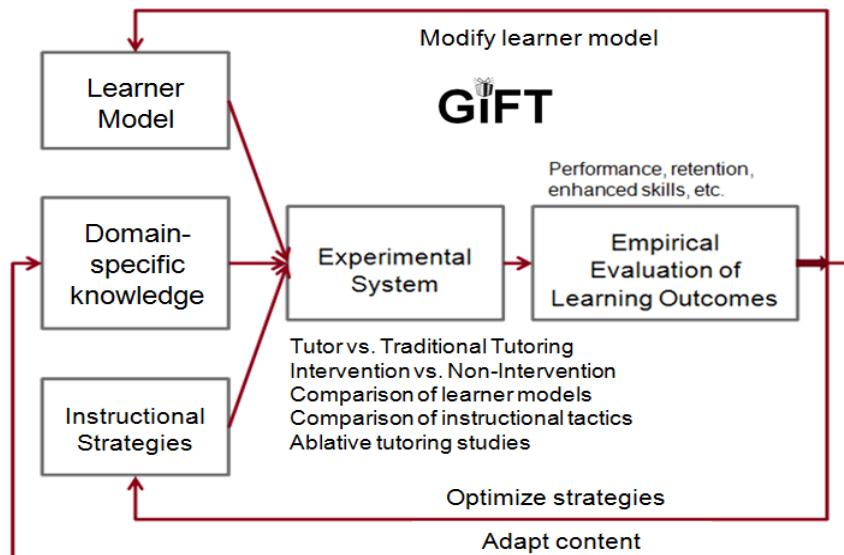
Figure P-2. GIFT Analysis Testbed Methodology

Figure P-2 illustrates an analysis testbed methodology being implemented in GIFT. This methodology was derived from Hanks, Pollack, and Cohen (1993) to allow manipulation of the learner model, instructional strategies, and domain-specific knowledge within GIFT, and support analysis of artificially-intelligent agents that influence the adaptive tutoring learning effect chain. In developing their testbed methodology, Hanks et al. reviewed four testbed implementations (Tileworld, the Michigan Intelligent Coordination Experiment [MICE], the Phoenix testbed, and Truckworld) for evaluating the performance of artificially intelligent agents. Although agents have changed substantially in complexity during the past 20–25 years, the methods to evaluate their performance have remained markedly similar.

The authors designed the GIFT analysis testbed based upon Cohen's assertion (Hanks et al., 1993) that testbeds have three critical roles related to the three phases of research. During the exploratory phase, agent behaviors need to be observed and classified in broad categories. This can be performed in an experimental environment. During the confirmatory phase, the testbed is needed to allow more strict characterizations of agent behavior to test specific hypotheses and compare methodologies. Finally, in order to generalize results, measurement and replication of conditions must be possible. Similarly, the

GIFT analysis methodology (Figure P-2) enables the comparison/contrast of ITS elements and assessment of their effect on learning outcomes (e.g., knowledge acquisition, skill acquisition, and retention).

How to Use This Book

This book is organized into four sections:

I. **The Influence of Affect, Engagement and Grit in Instructional Management**
II. **Metacognition and Self Regulated Learning**
III. **Natural Language and Discourse**
IV. **Instruction and Scaffolding**

Section I, *The Influence of Affect, Engagement and Grit in Instructional Management*, examines research, emerging concepts, and future directions for the *instructional management* of learner states by computer-based ITSs. Techniques, strategies, and tactics used by ITSs are reviewed with respect to their ability to enhance positive affect, moderate the influence of negative affect, promote engagement, and develop grit, also known as perseverance, as a desirable trait. Section II, *Metacognition and Self-Regulated Learning*, examines how metacognition (thinking about thinking) and self-regulated learning (self-initiated and self-managed instruction beyond the formal classroom environment) influence the design of ITSs. Section III, *Natural Language and Discourse*, reviews best practices of dialogue-based tutoring and their impact on ITS design. Section IV, *Instruction and Scaffolding*, focuses primarily on scaffolding and the Zone of Proximal Development as instructional strategies for equalizing the learner's domain competence with the challenge level of the domain content in order to maintain/promote engagement.

Chapter authors in each section were carefully selected for participation in this project based on their expertise in the field as ITS scientists, developers, and practitioners. *Design Recommendations for Intelligent Tutoring Systems: Volume 2 - Instructional Management* is intended to be a design resource as well as community research resource that can be of significant benefit as an educational guide for developing ITS scientists, a roadmap for ITS research opportunities, and a roadmap to the development and application of GIFT.

References

Aleven, V., McLaren, B., Roll, I. & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*, 101-128.

Anderson, J. R., Corbett, A. T., Koedinger, K. R. & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*, 167-207.

Anderson, L. W. & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of Educational Objectives: Complete edition*. New York : Longman.

Baker, R.S., D'Mello, S.K., Rodrigo, M.T. & Graesser, A.C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*, 223-241.

D'Mello, S. & Graesser, A.C. (2010). Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-adapted Interaction, 20*, 147-187.

D'Mello, S. K., Graesser, A. C. & King, B. (2010). Toward spoken human-computer tutorial dialogues. *Human Computer Interaction, 25*, 289-323.

Elson-Cook, M. (1993). Student modeling in intelligent tutoring systems. *Artificial Intelligence Review, 7*, 227-240.

Gagne, R. M. (1985). *The conditions of learning and theory of instruction* (4th ed.). New York: Holt, Rinehart & Winston.

Goldberg, B.S., Sottilare, R.A., Brawner, K.W. & Holden, H.K. (2011). Predicting Learner Engagement during Well-Defined and Ill-Defined Computer-Based Intercultural Interactions. In S. D'Mello, A. Graesser, , B. Schuller & J.-C. Martin (Eds.), *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII 2011) (Part 1: LNCS 6974)* (pp. 538-547). Berlin Heidelberg: Springer.

Graesser, A.C., Conley, M. & Olney, A. (2012). Intelligent tutoring systems. In K.R. Harris, S. Graham & T. Urdan (Eds.), *APA Educational Psychology Handbook: Vol. 3. Applications to Learning and Teaching* (pp. 451-473). Washington, DC: American Psychological Association.

Graesser, A. C., D'Mello, S. K., Hu. X., Cai, Z., Olney, A. & Morgan, B. (2012). AutoTutor. In P. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 169-187). Hershey, PA: IGI Global.

Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*, 104–137.

Hanks, S., Pollack, M.E. & Cohen, P.R. (1993). Benchmarks, Test Beds, Controlled Experimentation, and the Design of Agent Architectures. *AI Magazine*, *14* (4), 17-42.

Johnson, L. W. & Valente, A. (2008). Tactical language and culture training systems: Using artificial intelligence to teach foreign languages and cultures. In M. Goker & K. Haigh (Eds.), Proceedings of the Twentieth Conference on Innovative Applications of Artificial Intelligence (pp. 1632-1639). Menlo Park, CA: AAAI Press.

Krathwohl, D.R., Bloom, B.S. & Masia, B.B. (1964). *Taxonomy of Educational Objectives: Handbook II: Affective Domain*. New York: David McKay Co.

Lepper, M. R., Drake, M. & O'Donnell-Johnson, T. M. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds), *Scaffolding learner learning: Instructional approaches and issues* (pp. 108-144). New York: Brookline Books.

Litman, D. (2013). Speech and language processing for adaptive training. In P. Durlach & A. Lesgold (Eds.), *Adaptive technologies for training and education*. Cambridge, MA: Cambridge University Press.

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, *10*(1), 98–129.

Murray, T. (2003). An Overview of Intelligent Tutoring System Authoring Tools: Updated analysis of the state of the art. In Murray, T.; Blessing, S.; Ainsworth, S. (Eds.), *Authoring tools for advanced technology learning environments* (pp. 491-545). Berlin: Springer..

Merrill, D., Reiser, B, Ranney, M., and Trafton, J. (1992). Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. The Journal of the Learning Sciences, 2(3), 277-305

Nkambou, R., Mizoguchi, R. & Bourdeau, J. (2010). *Advances in intelligent tutoring systems*. Heidelberg: Springer.

Patil, A. S. & Abraham, A. (2010). Intelligent and Interactive Web-Based Tutoring System in Engineering Education: Reviews, Perspectives and Development. In F. Xhafa, S. Caballe, A. Abraham, T. Daradoumis & A. Juan Perez (Eds.), *Computational Intelligence for Technology Enhanced Learning. Studies in Computational Intelligence* (Vol 273, pp. 79-97). Berlin: Springer-Verlag.

Person, N. K., Kreuz, R. J., Zwaan, R. A. & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction, 13*(2), 161–188.

Picard, R. (2006). Building an Affective Learning Companion. Keynote address at the 8th International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan. Retrieved from http://www.its2006.org/ITS_keynote/ITS2006_01.pdf

Psotka, J. & Mutter, S.A. (1988). *Intelligent Tutoring Systems: Lessons Learned*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rus, V. & Graesser, A.C. (Eds.) (2009). The Question Generation Shared Task and Evaluation Challenge. Retrieved from http://www.questiongeneration.org/.

Simpson, E. (1972). The classification of educational objectives in the psychomotor domain: *The psychomotor domain*. Vol. 3. Washington, DC: Gryphon House.

Sleeman D. & J. S. Brown (Eds.) (1982). *Intelligent Tutoring Systems*. Orlando, Florida: Academic Press, Inc.

Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education, 12*(1), 40-62.

Sottilare, R. & Gilbert, S. (2011). Considerations for tutoring, cognitive modeling, authoring and interaction design in serious games. *Authoring Simulation and Game-based Intelligent Tutoring workshop at the Artificial Intelligence in Education Conference (AIED) 2011*, Auckland, New Zealand, June 2011.

Sottilare, R., Holden, H., Brawner, K. & Goldberg, B. (2011). Challenges and Emerging Concepts in the Development of Adaptive, Computer-based Tutoring Systems for Team Training. *Interservice/Industry Training Systems & Education Conference*, Orlando, Florida, December 2011.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). *The Generalized Intelligent Framework for Tutoring (GIFT)*. Orlando, FL: U.S. Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED).

Sottilare, R. (2012). Considerations in the development of an ontology for a Generalized Intelligent Framework for Tutoring. *International Defense & Homeland Security Simulation Workshop* in Proceedings of the I3M Conference. Vienna, Austria, September 2012.

Sottilare, R., Ragusa, C., Hoffman, M. & Goldberg, B. (2013). Characterizing an adaptive tutoring learning effect chain for individual and team tutoring. In Proceedings of the *Interservice/Industry Training Simulation & Education Conference*, Orlando, Florida, December 2013.

VanLehn, K. (2006) The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education. 16*(3), 227-265.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221.

Woolf, B.P. (2009). *Building intelligent interactive tutors*. Burlington, MA: Morgan Kaufmann Publishers.

PROLOGUE

Benjamin D. Nye[1], Robert A. Sottilare[2], Charles Ragusa[3] and Michael Hoffman[3]

University of Memphis[1]

U.S. Army Research Laboratory [2]

Dignitas Technologies, LLC[3]

Defining Instructional Challenges, Strategies, and Tactics for Adaptive Intelligent Tutoring Systems

Background

Instructional strategies play a critical role in intelligent tutoring systems (ITSs), human one-to-one tutoring, and traditional classroom instruction. They are the mechanisms within an ITS that determine the optimal course of action to improve student learning. However, the definition and purpose of instructional strategies lacks clear consensus. Despite this ontological roadblock, the concept of an "instructional strategy" has strong theoretical and practical implications for ITS designs and is especially relevant to generalized tutoring architectures. This prologue provides four arguments in support of the goal of a standard purpose and definitions for instructional strategies and tactics. First, we review the purpose of instructional strategies and tactics by reviewing their theoretical underpinnings in the literature and the importance of these concepts for ITSs. Second, we put forth a set of standard definitions for adaptive instructional strategy concepts, which include categories of strategies and tactics. Finally, we examine how these standard concepts might be represented and implemented in GIFT, a tutoring architecture that is attempting to capture standards for authoring, automated instruction, and evaluation of the effects of ITS technologies. Finally, the future of instructional strategies and tactics for adaptive ITSs are considered from the perspective of ITS researchers, developers, authors, and end-users.

Figure 1 shows an archetypal breakdown of ITS components plus interaction with the student or learner. This diagram has been expanded on at length by texts such as *Building Intelligent Tutoring Systems* (Woolf, 2009), but dates back to at least the 1980s (e.g., *Foundations of Intelligent Tutoring Systems*; Polson & Richardson, 1988). Often, these diagrams only note the existence of the four components and imply that the student interacts through the communication interface. The arrows in Figure 1 indicate typical pathways that information moves between modules, but do not specify information types or formats. In general, an ITS has four primary functions at runtime: a communication interface, a domain model, pedagogical model, and a student or learner model. The communication interface receives student input, and presents feedback to the student. The domain model contains information that is specific to the content that the ITS teaches. The student model classifies states of the learner (e.g., cognitive, affective, performance) to determine the student's progress toward the mastery of presented concepts). The relationships between these models vary between ITSs, but these roles are quite common.
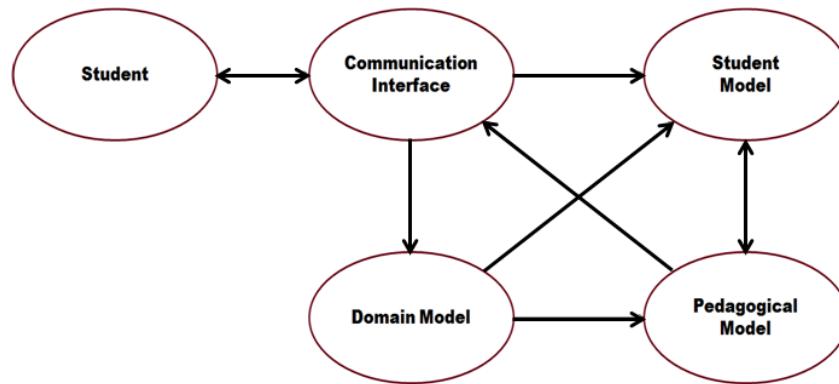

Figure 1. Archetypal Four-Component Tutoring System Design (plus Student).

The pedagogical model handles the actual "tutoring" aspects of the tutoring system. A pedagogical module determines what information to present to the student and what information is needed from the student to determine the student's knowledge and skill, as compared to the expected knowledge and skill at any point in the tutoring process. In many ITSs, expected knowledge and skills are used as standards that are part of the domain model and are called the *expert model* or *ideal student model*.

In a typical system, the domain and student models provide information on progress and student state, respectively, to the pedagogical model. This information is used to make decisions on which actions to take next. After the pedagogical module has decided on an action (e.g., ask a question), the communication module determines how to present this information or modify the environment to reflect recommended changes. The effectiveness of the ITS can only be as good as the strategies within the pedagogical model. Some research has indicated that using machine learning techniques to fine-tune these strategies might increase the learning gains realized by an ITS significantly (Chi, VanLehn & Litman, 2010). However, the definition and nature of these instructional strategies is not easy to pin down, due to significant differences in how systems and scholars conceptualize and implement instructional strategies.

An "instructional strategy" could refer to a whole learning theory (e.g., direct instruction vs. a constructivist approach), a specific heuristic (rule of thumb) used by a teacher, a learning principle extracted from studies of cognition, or a set of rules within an ITS. The concept of an instructional strategy predates ITSs and has been progressively extended to accommodate not only different representations, but entirely different levels of analysis when considering how to select pedagogically useful behaviors. Strategies may also exist at multiple levels, such as the macro-adaptive level (e.g., selecting course units or assigning instructional tasks) or the micro-adaptive level (e.g., helping with assigned tasks or optimizing presentation and communication). Other terms sometimes used interchangeably with the term *instructional strategy* are *pedagogic strategy*, *pedagogical strategy*, *teaching strategy*, *instructional tactic, instructional technique*, or *instructional principle*.

Whatever we choose to call them, instructional strategies are important because they form the foundation for critical decisions by the ITS to provide content or feedback, change the challenge level of a scenario, or decide when to move forward to the next concept in a lesson. An ITS uses strategies during instruction to adapt based on the student's learning needs, as identified by the student's performance, states, and traits. We envision this will also be the case for teams as ITS technologies are extended to collaborative learning environments.

Expert Viewpoints on Instructional Strategies

A meeting of ITS experts was convened in July 2013 at the University of Memphis to discuss instructional strategies and make recommendations for the design of GIFT's pedagogical module. This advisory board consisted of leading academic and government scientists in the field. Many of these experts have authored chapters within this book. Despite the challenge of unambiguously defining instructional strategies, the group had a clear consensus that instructional strategies were both useful and important. The group noted that instructional strategies add value to the design, evaluation, and improvement of ITSs in at least three ways: 1) by representing an expert model of pedagogy; 2) by separating pedagogical behavior from domain knowledge; and 3) by grouping ITSs into different categories, which helps us to study and compare their effectiveness in a variety of tutoring domains.

First, expert models may be developed by an expert teacher who has deep understanding of the domain (Mitrovic, Martin & Suraweera, 2007) or by learning these strategies from data sets collected from human tutoring sessions (Graesser, D'Mello et al., 2012). Second, the separation of pedagogical behavior from domain knowledge presents distinct advantages in generalized tutoring systems where the goal is to reuse

instructional strategies for new domains (Sottilare, Goldberg, Brawner & Holden, 2012). Finally, given cognitive processes vary across different learning tasks and that the human mind is constrained by limited working memory, certain instructional approaches may be more effective than others for different problems and people (Koedinger & Corbett, 2006). Categorizing ITSs by their instructional strategies can be used to help evaluate learning effects across different systems, domains, and student populations.

Experts in the field also noted differences in the goal of instructional strategies, their implementation, and scope. The next section discusses these viewpoints and attempts to identify categories of instructional strategies that are based on these differences.

Instructional Goals for Strategies

The experts had a strong consensus that instructional strategies were intended to guide the learner toward a set of instructional goals, though they did not necessarily agree on the nature of those goals. While some experts focused on domain learning outcomes as goals, others believed that instructional strategies implied a focus on longer-term goals such as real-life applications, transfer learning, or self-regulated learning. Still other experts saw instructional strategies as a methodology for customizing goals for each user or curriculum.

These offer three different perspectives on the goals for an instructional strategy. In the first perspective, an instructional strategy makes decisions that are intended to improve the learner's knowledge of the domain trained by the ITS. This is probably the most common view of an instructional strategy. ITSs often focus on helping a student learn a particular topic or set of domain skills that align with curriculum goals. In the second perspective, an instructional strategy focuses primarily on longer-term learning outcomes. These could include retention, transfer of learning to other operational contexts, learning-to-learn, or metacognition where a student builds skills that helps make learning more efficient (Azevedo & Cromley, 2004; Chi & VanLehn, 2007). In the final case, the ITS acts as a framework for curriculum designers but does not inherently assume learning goals. This allows a learner, group of learners, or teacher to actively construct personalized goals, such as through self-regulated learning. For example, Betty's Brain focuses on causal systems such as ecology and allows the students to select and explore different causal relationships (Biswas, Segedy & Kinnebrew, 2013).

These different views of goals demonstrate that ITSs have been built with a range of goals in mind, ranging from a sharp focus on specific domain concepts to allowing ad-hoc learner-defined goals that the strategies work to support. Despite different views of the best goals for strategies, experts appear to have a consensus that strategies select ITS behavior that guides a learner toward one or more instructional goals. However, these goals might not need to be the same for different systems or even for different users of the same system.

Discrete vs. Continuous Representations for Strategies

Another viewpoint on instructional strategies contrasts discrete and continuous representations of instructional strategies. If instructional strategies drive a decision-making process that guides learners toward particular outcomes, then strategies must somehow be represented as functions or processes. A "conditional" system implies a set of rules or boundaries, which are used to determine tutoring behavior. Considering instructional strategies as a "policy" implies that they might be framed in terms of a Markov Decision Process (MDP) that maximizes learning over a given time horizon (Puterman, 2009). Finally, considering instructional strategies as "planning" to reach a particular goal could imply that the purpose of an ITS is to perform "path planning" that guides the learner toward a particular learning state, while

minimizing certain costs, such as study time. Fundamentally, these are just different representations of the problem, which should yield similar tutoring behavior if implemented optimally.

However, from the standpoint of implementing an ITS, they can imply significantly different designs. One difference between these perspectives is the difference between considering instructional strategies from a discrete versus a continuous perspective. Discrete, state-based systems are represented strongly among ITSs, with both constraint-based and production rule systems relying on discrete conditions to select actions. GIFT currently uses a rule-based decision process that is driven by transitions between discrete states (Sottilare et al., 2012). However, continuous representations of strategies are equally valid. Rather than breaking the tutor's knowledge into discrete states or conditions (e.g., right answer vs. wrong answer), a tutoring system can instead calculate the expected utility of actions based on continuous features (i.e., a multi-attribute utility).

In a discrete space, such as one based on conditional rules, knowledge from the student model and domain are used to determine discrete states that map to a particular optimal action set. For example, an ITS might employ a simple strategy that selects between three possible actions: a Concept Map Task (e.g., drawing semantic links between concepts), a Tutorial Dialog (e.g., natural language dialog with a tutoring avatar), and No Action. Two rules are evaluated, one to determine if the learner has low knowledge (Low Knowledge) and one to determine if the learner learns best from verbal explanations (Verbal Learner). By evaluating the combinations of these rules, the appropriate intervention can be selected. By comparison, a hypothetical utility function can calculate continuous equivalents to the rules in the discrete version: Knowledge Level and Verbal Learning. By calculating the utility of each action, the strategy can select the best action or no action, if no action has a positive utility. For this small example, the continuous representation can be easily reduced to the discrete version by stating threshold rules for the continuous inputs. Even for more complicated strategies, discrete equivalents can produce the same decisions, so long as the actions are discrete. So then, this difference does not fundamentally define the behavior or quality of a tutoring system. However, these choices may significantly affect the effort to represent certain strategies.

Some of our experts noted that optimal instructional behavior often balances multiple interacting or competing facets. Representing instructional strategies as a continuous field of action utilities captures this intuition more naturally than a conditional system. A rule-based system must segment the state space to define the action(s) that should occur in each case. By comparison, a continuous system can weigh which actions are better for a given state and pick the best one(s). This does not mean that continuous representations (e.g., utility-based agents, certain MDPs) are better than discrete ones. For example, conditional systems allow a high degree of control and interpretability over strategy decision making. By comparison, determining the higher utility value for two possible actions is seldom easy for a human to interpret. These differences probably indicate that continuous or discrete representations for strategies are different enough to offer advantages for different domains or scopes of learning goals. Mixed representations (i.e., hybrid systems) are a third option that may allow the greatest flexibility, though at the cost of increased complexity for creating a system.

Scope of Strategies: Domain-Independence and Domain-Dependence

A related debate muddies the distinction between domain-independent strategies versus domain-dependent strategies. The definition of a "domain independent" instructional strategy is ambiguous. If we take instructional strategies as an approach to move a learner toward some learning goals, it is not entirely clear which strategies could move every learner toward any arbitrary learning goal. Moreover, even for strategies that could apply to any learning goal in any domain, it would defy reason to expect that the same strategy would work equally well for all domains and learning goals. Should a strategy be consid-

ered domain-independent if it works twice as well in one domain compared to another? Unfortunately, without clear cut-and-dry standards for what makes a strategy domain-independent, this question does not have a direct answer.

To get a better handle on this distinction, it is important to pin down what domain-independence means for a strategy. There are two facets to domain-independence: domain neutrality and effectiveness. By domain neutrality, we mean that a domain-independent strategy must not reference or rely upon any domain-specific information. It may be able to receive domain-specific information and process it using its domain-independent mechanisms, but it cannot contain any assumptions specific to a domain. Without this requirement, a strategy could explicitly rely on domain-specific features. This must be considered a minimum requirement for a domain-independent strategy. However, depending on how a domain is defined, the space of strategies with no domain assumptions may be quite small. This implies that instead of a two-category system (independent vs. dependent), strategies might instead be considered in terms of the set of domains where they apply. This perspective is considered in more detail later in this section.

The effectiveness of a strategy for a domain might also be considered. Effectiveness means that the strategy must bring the learner closer to instructional goals. Obviously, a domain-independent instructional strategy should be useful for instructional purposes across different domains. However, there are no clear-cut standards for establishing overall effectiveness of different instructional strategies. Attempts to classify strategies to improve evaluation, such as the Framework for Instructional Technology (FIT) model, have been proposed, but more empirical evaluations are needed to understand the interaction of strategies and different domains (Durlach, 2012). Despite this, there could still be long-term value in considering domain-dependence in terms of the relative effectiveness of strategies for different domains.

These perspectives on domain-dependence and independence can be stated more formally. Assume that D is the set of all domains taught by ITSs, where each specific domain d (where d is a member of D) has a set of G_d possible instructional goals. Second, assume that s is an instructional strategy and D_s is the set of domains where that strategy can be implemented, regardless of effectiveness. Next, assume that $E(s, g_d)$ represents some evaluation function for the effectiveness of a strategy s for the learning goal g_d, where $E(s, g_d) > 0$ indicates that the strategy brings a student closer to that goal, $E(s, g_d) = 0$ indicates that it is ineffective and $E(s, g_d) < 0$ indicates that it actually hinders learning (e.g., introduces misconceptions).

Table 1 shows the implications of different perspectives on domain independence versus domain dependence. In the first case, strategies must be categorized as either domain-independent or domain-specific. A domain-independent strategy can be used for all the domains of interest (D). A domain-specific strategy is effective for only one domain, so the set ($|D_S|$) has exactly one member. This has the clear limitation that the majority of strategies might be applicable to multiple domains, but not all domains. This category scheme does not give a way to represent that scenario. The "Set of Domains" case accommodates these cases that fall through the cracks. While it has the same definition of domain-independence, a domain-dependent strategy is classified as one that cannot be applied to one or more domains. Under this representation, domain-independence is merely a special case of a strategy being applicable to multiple domains. This approach is also compatible with different ways to classify domains, because it relies on listing out the domains where each strategy can be used.

Table 1. Different Types of Domain Dependence and Independence.

Viewpoint	Domain-Independent	Domain-Dependent
Domain-Independent vs. Domain-Specific Categories	$D_s = D$	$|D_s| = 1$
Set of Domains for Each Strategy s	$D_S = D$	$D \setminus D_S \mathrel{!=} \varnothing$
Relative Effectiveness of Strategy s for Domains	$E(s,\ g_d) > 0 \qquad \forall\ d \in D$ $\exists\ g_d \in G_d$	$E(s,\ g_d) \leq 0 \qquad \exists\ d \in D$ $\forall\ g_d \in G_d$

In the final viewpoint, a domain-independent instructional strategy must be at least marginally effective in all domains. This would require $E(s,\ g_d) > 0$ for at least one meaningful learning goal g_d in order to be considered an effective strategy for the domain d. A domain-dependent strategy would be the converse of this: for at least one domain, it is ineffective or hinders all instructional goals. This standard is obviously a very low bar to cross: the strategy would only need to be better than nothing for a single learning goal. Obviously, more stringent conditions might be applied, such as being more effective than some threshold (i.e., $E(s,\ g_d) > c$) or being effective for a larger number of learning goals within a domain. A domain-independent strategy might be defined as one that satisfies this sort of condition for all or most domains of pedagogical interest.

However, this raises the question: what have we gained by converting our hypothetical continuous measure of effectiveness into a Boolean "domain-independent" categorization? There is no theoretical benefit for representing it along these lines. However, from the standpoint of ITS authors, a simple classification is far easier to evaluate than a large set of estimates of effectiveness. With that said, similar or better authoring intuitions might arise from specific measures of the effectiveness of a strategy for a particular domain or from the average effectiveness of a strategy across many domains. Unfortunately, evaluation of the relative effectiveness of strategies across different domains is sparse. This probably makes a graded or continuous approach to considering effectiveness infeasible in the near term.

So far, we have identified the range of terms, definitions, and contexts for the use of instructional strategies within the literature and we have solicited the opinions of experts in the ITS field. In the following section begin to define and support a set of terms and definitions for instructional strategies within adaptive ITSs.

Defining a Hierarchical Concept of Instructional Strategies

As we have examined the dimensions of instructional strategies, it is now time to define a relationship between these dimensions and boundaries of the term *instructional strategies*. Throughout this book, the reader is likely to see wide array terms to imply instructional strategy: *pedagogic strategy*, *pedagogical strategy*, *teaching strategy*, *instructional tactic, instructional technique*, or *instructional principle*. Since we are attempting to establish standards for GIFT and its user community, we pose the following hierarchical concept of instructional strategies.

For our purposes, instructional strategies are "plans, recommendations, and processes provided by the ITS to bring the student closer to the instructional goals, which are generally initiated by instructional challenges." *Instructional goals* include, but may not be limited to, enhanced learning (knowledge and skill acquisition), accelerated learning, enhanced performance (application of knowledge and skill), enhanced retention, enhanced engagement (increased opportunity for learning), and enhanced motivation (increased

potential to overcome difficulties and persist in learning). Effective ITSs are effective because they influence student progress toward these goals.

The adaptive tutoring learning effect chain, a model for learner-ITS interaction, is shown in Figure 2. Instructional strategies are considered to be largely domain-independent and may be either macro-adaptive (pre-instructional) or micro-adaptive (during instruction). Recommendations of courses or lessons based on the student's previous training are an example of a macro-adaptive strategy, while a recommendation to ask the student a question to assess learning is a micro-adaptive strategy based on current performance. A hierarchical relationship exists between instructional strategies and instructional tactics. *Instructional tactics* are "actions taken by the ITS in response to strategies (plans or recommendations)." Tactics are domain-specific actions. For example, the selection and presentation of a context-relevant question by the tutor is an instructional tactic.
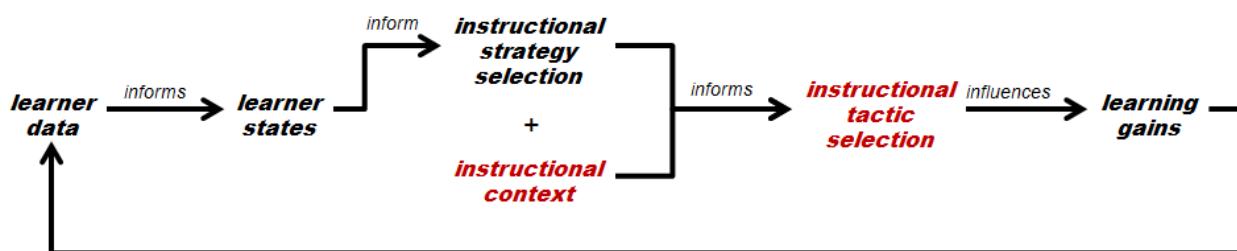

Figure 2. Adaptive Tutoring Learning Effect Chain (Sottilare, Ragusa, Hoffman & Goldberg, 2013).

Within the context of instruction, the ITS adapts instruction in response to either an observation of change in the learner's[1] state (e.g., performance, cognitive, affective) or the learning environment (e.g., game, webpage, tutor-user interface), which is sufficiently significant to trigger a decision by the ITS. For our purposes, *sufficiently significant* implies that an *instructional challenge* has been identified by the tutor. For example, the learner makes an error in attempting to solve a problem. The decision is for the ITS to intervene now or wait until another mistake is made. If the ITS assesses the error is significant enough to warrant an intervention now, the next decision is which instructional strategy to implement: feedback, review of content, or content modification. Once a strategy is selected, the specific action to select and present/modify information is a tactic.

It is also important to note what instructional strategies are not. Instructional strategies are not *learning strategies*. Unlike an instructional strategy, which is initiated by the tutor, a learning strategy is owned by the student and is the student's approach to understanding information, building/rebuilding mental models, and using these models to solve problems. Good instructional strategies should reinforce effective learning outcomes and may do so by reinforcing proven learning strategies. So, instructional strategies and learning strategies are related, but they are different.

Instructional strategies are also not *educational philosophies* or theories. Constructivism is an educational philosophy in which learners are encouraged to "work together and support each other as they use a variety of tools and information sources in their guided pursuit of learning goals and problem-solving activities" (Perkins, 1991). This is relevant to instruction since learners construct knowledge from information generated from previous experiences. While educational philosophies like constructivism can and should drive the design of ITSs, they illustrate generalized goals and are not instructional strategies.

---

[1]The terms *learner*, *student*, and *user* are used interchangeably within this book. However, *users* may in some instances be more expansive and include researchers, developers, and designers in addition to learners/students.

In the following section, the approach used by the GIFT architecture for separating domain-specific and domain-independent instructional decision-making is discussed. GIFT, which distinguishes between instructional strategies (domain-independent) and tactics (domain-specific) offers a useful case study for considering this topic.

Implementation of Strategies and Tactics in GIFT

In early versions of GIFT, the pedagogical model relied on rules that used observed changes in learner state to trigger abstract instructional strategies such as instructional interventions or scenario adaptations (e.g., feedback or changes in difficulty, respectively). The abstract pedagogical requests were then sent to the domain module where they were translated into concrete strategy implementations (tactics), relevant to the current learning context. Message flow between the GIFT modules is illustrated in Figure 3.
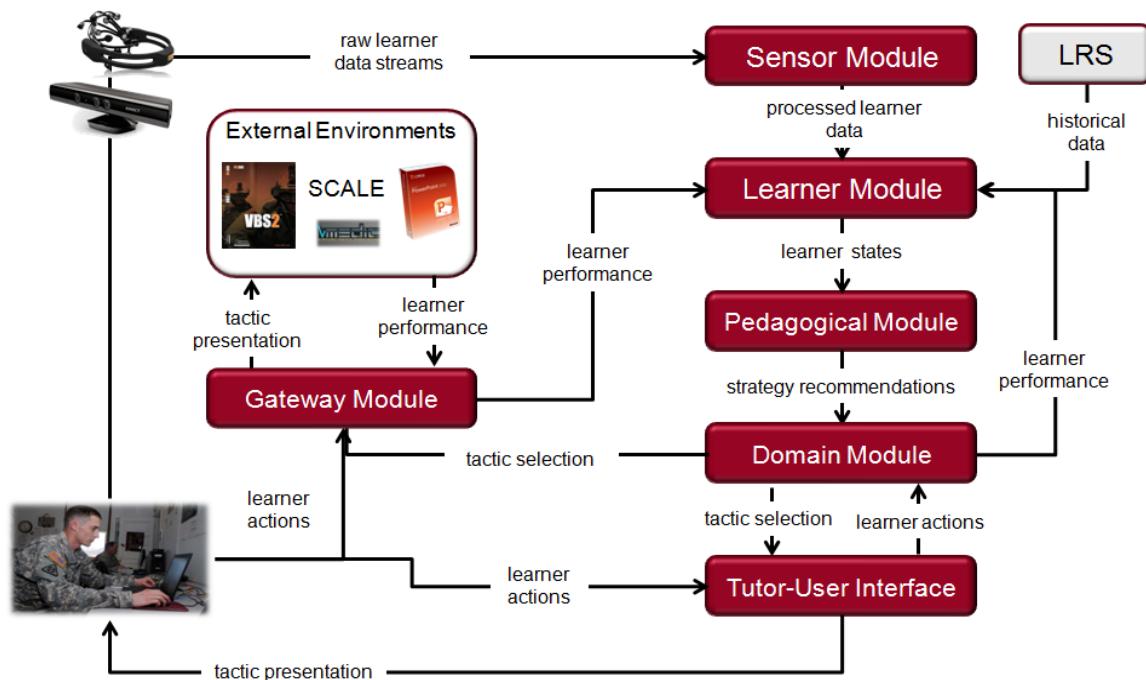
Figure 3. Real-Time Micro-Adaptive Strategies in GIFT.

In recent development by the GIFT team, this same basic framework has been extended to support the engine for Managing Adaptive Pedagogy (eMAP). The first iteration of eMAP (Sottilare et al., 2013) allowed inclusion of dynamic branching through Merrill's (1983) Component Display theory presentation quadrants based on metadata tags for content. At each Merrill's branch point the eMAP used the current quadrant (i.e., rules, examples, recall, and practice) together with the current learner state(s) to identify the preferred metadata attributes of the next quadrant. These attributes were then compared to the metadata attributes of the next-quadrant choices and the best match selected, resulting in presentation of the associated content to the learner.

In the current development cycle, the eMAP is being further extended to support more advanced flow through course content, including support for mastery learning, as illustrated in Figure 4. In this implementation, GIFT will maintain a hierarchical representation of the course concepts (created for each course by the course author) and will include one or more (preferably more) units of content covering

each node in the concept tree. Coverage of non-leaf nodes may be explicit or inferred by roll up of child nodes. Metadata tags on content will be extended to include concept names, thus providing linkage back to the concepts in the concept hierarchy. Finally, survey (quiz) questions will be authored and tagged to be used as checks on learning, again linking questions back to one or more concepts in the hierarchy.

Within a lesson, learners will proceed through Merrill's quadrants as before, but now checks on learning will serve as gates. For example, in the recall quadrant, GIFT will use tagged questions to create an ad-hoc concept survey. User responses to the survey will be scored and used to assess learning. Demonstrating mastery of concepts will allow the learner to advance either to guided practice and reflection, or to the next lesson. Failure to demonstrate mastery will route the user through a remedial path. Similarly, success in the guided practice is required for advancement, where as failure results in remediation, as shown in Figure 4.
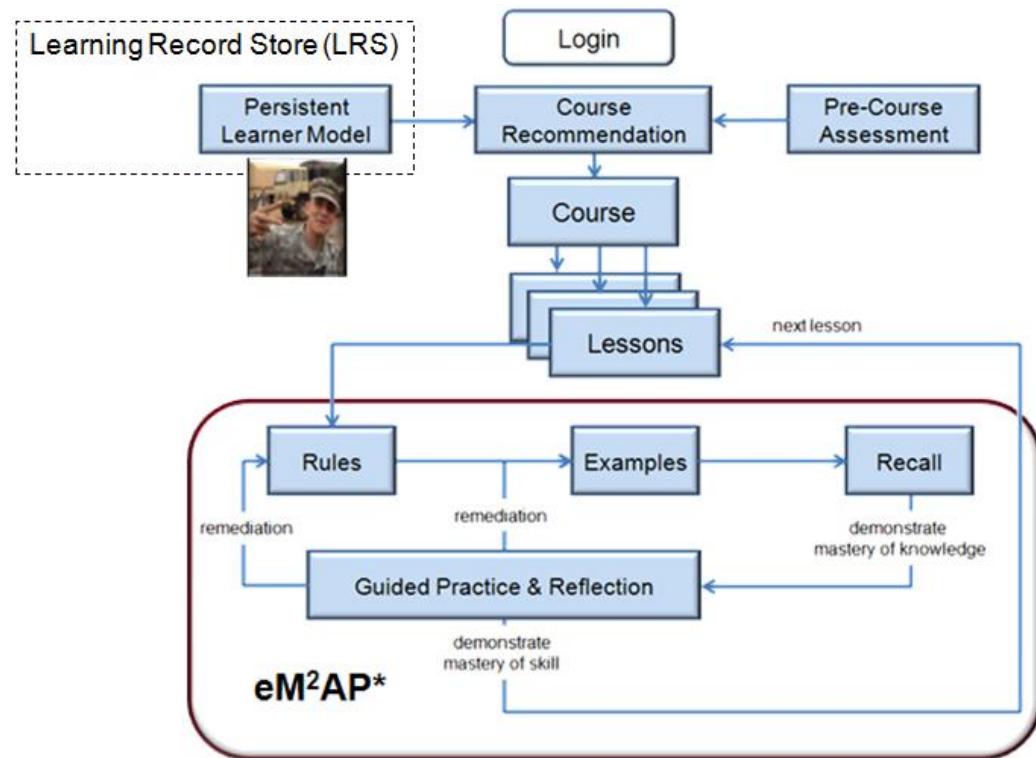
Figure 4. Macro-Adaptive Strategies in GIFT using eMAP.

Future Capabilities of Instructional Strategies

Instructional strategies currently support tutoring system macro-adaption (e.g., task selection) and micro-adaption (e.g., step-based support for learning). In the future, as ITSs integrate with persistent learning systems, we should expect instructional strategies to play a significant role at the curriculum-planning level: personal learning assistants to help students find courses that support their lifelong learning needs. We should also expect the growth of ITSs that contain a large variety of instructional strategies, switched dynamically to target different types of learners. These might dispatch tutoring to other tutoring systems or even to other humans, creating a cybernetic tutoring experience (hybrids of ITS and human tutoring). For example, "teacher in the loop" tutoring systems would offer a powerful hybrid that maximizes the effective traits of both humans and computers. Collaborative use of ITSs is a particular area that may

require new strategies, such as team-based learning. ITSs have typically focused on one-on-one strategies, so this could be a genuine paradigm shift. GIFT, in particular, is trying to play a significant role to integrate different tutoring systems and collaborative uses into a unified platform.

Authoring is another major area for tutoring systems. A serious challenge for authoring tool design is managing the tradeoff between flexibility and simplicity. Some authoring tools provide advanced capabilities, to the point of supporting Turing-complete computing. For example, AutoTutor Authoring Tools (ASAT) allows authors to write production rules to power tutoring dialog (Graesser et al., 2004). Others provide highly constrained authoring capabilities, such as the Example Tracing authoring in the Cognitive Tutor Authoring Tools (CTAT; Aleven, McLaren, Sewall & Koedinger, 2009). Neither approach is perfect. A highly flexible tool leads to authoring complexity and often a longer learning curve. On the flip side, a highly constrained authoring tool forces the author to use a limited set of instructional strategies.

Instructional strategies may someday be used as templates for authoring a tutoring system to help balance these tradeoffs. Each instructional strategy requires certain authoring input (e.g., hint statements) and supporting information (e.g., an affect-sensitive strategy needs an affect classifier). Templates or other authoring scaffolds for strategy-specific authoring could be designed and maybe someday generated automatically. These templates could constrain authors to the content that is most important to that strategy. A general authoring system that uses strategy-specific templates for authoring would offer an effective balance of power and simplicity. With that said, work on categorizing instructional strategies and identifying templates is needed before this is possible. The next major focus for GIFT will be to explore authoring tools for ITS.

Finally, instructional strategies can help to evaluate tutoring systems. As learning technologies become ubiquitous, the community must increasingly focus on which strategies work best for different contexts, populations, and cultures (Blanchard, 2012). Instructional strategies have a long track record for classifying instructional behavior. This paradigm can also be used to classify the behavior of tutoring systems. By classifying systems based on the strategies they use, these evaluations can be aggregated into larger meta-analyses to determine their value for different learners and domains. Strategy classifications could also play an important role in rating and recommender systems that support lifelong learning. They could aid institutions and learners in selecting the right learning technologies for their needs. In this way, instructional strategies can benefit developers, authors, and end-users working with ITSs.

References

Aleven, V., McLaren, B. M., Sewall, J. & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. International Journal of Artificial Intelligence in Education, 19(2), 105-154.

Azevedo, R. & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? Journal of educational psychology, 96(3), 523.

Biswas, G., Segedy, J. R. & Kinnebrew, J. S. (2013). Smart Open-Ended Learning Environments that support Learners Cognitive and Metacognitive Processes. Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data, LNCS 7947 (pp. 303-310) New York, NY: Springer.

Blanchard, E. G. (2012, January). On the WEIRD Nature of ITS/AIED Conferences. In Intelligent Tutoring Systems (pp. 280-285). Springer Berlin Heidelberg.

Chi, M. & VanLehn, K. (2007, June). Accelerated Future Learning via Explicit Instruction of a Problem Solving Strategy. In Proceedings of Artificial Intelligence in Education (AIED) 2007 (pp. 409-416). IOS Press.

Chi, M., VanLehn, K. & Litman, D. (2010). Do micro-level tutorial decisions matter: Applying reinforcement learning to induce pedagogical tutorial tactics. In Proceedings of Intelligent Tutoring Systems (ITS) 2010 (pp. 224-234). Springer: New York, NY.

Durlach, P. J. (2012). Vanilla, Chocolate, or Chunky Monkey: Flavors of Adaptation in Instructional Technology. Proceedings of iFest 2012, Retrieved October 3, 2013 from www.adlnet.gov/wp-content/uploads/2012/08/Durlach_Adaption_in_IT_iFest-2012.pdf

Graesser, A. C., D'Mello, S. K., Hu. X., Cai, Z., Olney, A. & Morgan, B. (2012). AutoTutor. In P. McCarthy and C. Boonthum-Denecke (Eds.), Applied natural language processing: Identification, investigation, and resolution (pp. 169-187). Hershey, PA: IGI Global.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A. & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. Behavior Research Methods, Instruments & Computers, 36(2), 180-192.

Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. Educational Psychologist, 41(2), 75-86.

Koedinger, K. R. & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. The Cambridge Handbook of the Learning Sciences (pp. 61-78). Cambridge: Cambridge University Press.

Merrill, M. D. (1983). Component Display Theory. In C. M. Reigeluth (Ed). Instructional-design theories and models: An overview of their current status, pp.279-333. Hillsdale, NJ: Lawrence Erlbaum Associates

Mitrovic, A., Martin, B. & Suraweera, P. (2007). Intelligent Tutors for All: The Constraint-Based Approach. IEEE Intelligent Systems, 22(4), 38-45.

Perkins, D.N. (1991). Technology meets constructivism: Do they make a marriage? Educational Technology, 31, 18–23.

Polson, M.C. & Richardson, J.J. (1988). Foundations of Intelligent Tutoring Systems. (pp.1-19).

Puterman, M. L. (2009). Markov decision processes: discrete stochastic dynamic programming (Vol. 414). Hoboken, NY: John Wiley and Sons.

Sottilare, R. A., Goldberg, B., Ragusa, C. & Hoffman, M. (2013). Characterizing an Adaptive Tutoring Learning Effect Chain for Individual and Team Tutoring. In The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (Vol. 2013, No. 1). National Training Systems Association. Orlando, Florida, December, 2013.

Sottilare, R. A., Goldberg, B. S., Brawner, K. W. & Holden, H. K. (2012). A Modular Framework to Support the Authoring and Assessment of Adaptive Computer-Based Tutoring Systems (CBTS). In The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (Vol. 2012, No. 1). National Training Systems Association. Orlando, Florida, December, 2013.

Woolf, B. P. (2009). Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann.

Affect, Engagement and Grit in Instructional Management

R. Sottilare, Ed.

CHAPTER 1 – Thoughts on the Instructional Management of Affect, Engagement, and Grit

Robert A. Sottilare
U.S. Army Research Laboratory

Introduction

This first section of this book examines research, emerging concepts, and future directions for the *instructional management* of learner states by computer-based ITSs. Specifically, this section discusses techniques, strategies, and tactics used by ITSs to enhance positive affect, moderate the effects of negative affect, promote engagement, and develop grit as a desirable trait. All are discussed in terms of their effect on learning (knowledge and skill development), performance, and the ability to solve problems. For purposes of our discussion, affect states range in persistence from long to short as personality, mood, and emotions. Engagement is discussed as a necessary precursor to learning and grit is used interchangeably with perseverance or the ability of the learner to persist over long periods of time to progress toward goals in the face of significant challenges.

Enhancing the Intelligence of Intelligent Tutoring Systems

The ability to adapt is signpost of intelligence. As we strive to develop more adaptive ITSs, a key challenge is to enhance and optimize the decision making of tutoring systems. A goal for the design of ITSs is to fully automate the management of instruction so computers can guide one-to-one (individual) and one-to-many (team) tutoring efficiently and effectively. As discussed in *Design Recommendations for Intelligent Tutoring Systems: Volume 1 - Learner Modeling*, efficient tutoring will be largely dependent upon what the tutor "knows" about the learner so this knowledge can be used by the ITS to inform instructional decisions. Effective tutoring will be determined by optimizing instructional decisions to keep the learner in a positive affective state, engaged in the learning process and motivated to persist in the face of difficult and challenging learning concepts and conditions to get the best learning outcomes possible.

The chapters in this section focus on techniques, strategies, and tactics to enhance learning and performance. While not all concur with this approach, we have adopted the following instructional management taxonomy for the development of GIFT, a tutoring architecture to support automated authoring, automated instruction, and effect analysis. Instructional techniques are considered to be best practices for learning broadly applied in instructional systems (including, but not limited to, ITSs). Instructional techniques evolve over time based on lessons-learned and observed effect, and are applied largely without respect to who the learner is, what the domain being instructed is, and the specific instructional context. In other words techniques are learner- and domain-independent. Techniques only require information about learner performance (e.g., errors) and treat each learner the same. Strategies, on the other hand, are learner-dependent, but domain-independent. They can be considered *plans for future actions*. Actions taken by the ITS are tactics. Tactics are both learner-dependent and domain-dependent.

Managing Learner States and Guiding Instruction

The seven succeeding chapters in this section highlight ongoing areas of research related to techniques, strategies, and tactics used by ITSs to manage affect, engagement, and grit.

Chapter 2, by Sottilare, DeFalco, and Connor, provides a review of the literature related to generalizable instructional techniques and specific strategies for moderating affect, enhancing engagement, and assessing/developing grit or perseverance as a desirable trait. This chapter forms the basis of a literature review of affect, engagement, and grit management strategies. The literature review on affect primarily focuses in three areas: developing the emotional intelligence of ITSs to improve their capability to assess and optimally manage learner affect; modeling the response of expert human tutors to learner affect in order to transfer desirable behaviors, trait, and principles to computer-based tutors to improve learner-tutor interaction and the development of trust between learner and ITS; and finally, assessing one of the few implementations of Vygotsky's Zone of Proximal Development (ZPD) with ITSs. The primary contribution of this chapter is the analysis of the effect of ITS cognitive design principles on affect.

In chapter 3, D'Mello, Blanchard, Baker, Ocumpaugh, and Brawne discuss affect-sensitive strategies. Learner affect ranges in duration from momentary expressions (e.g., aha and eureka moments) to more enduring attitudes that can moderate learner decisions, levels of engagement, and motivation. D'Mello et al. note that affect indirectly influences learning outcomes (e.g., knowledge acquisition) by modulating cognitive processes during instruction. The primary contribution of this chapter is an exposition of six case studies, each featuring a unique, tried and tested, affect-sensitive instructional strategy.

Chapter 4, by DeFalco, Baker, and D'Mello, discusses several types of interventions mentioned in the literature to combat disengaged behaviors in online learners and examines the potential of adaptive interventions in other contexts to reduce or eliminate behavioral disengagement. It is generally accepted that GIFT and other tutoring delivery systems will provide on-demand tutoring at the point-of-need of the learner via service-oriented architectures (online learning resources). The major contribution of this chapter is evaluation of interventions that potentially can be incorporated into the GIFT framework for broad use by the tutoring community. These interventions are intended to promote engaged behaviors conducive to focused, deep learning, through reducing disengaged behaviors that are not conducive to focused, deep learning. Desirable engaged behaviors include, but are not limited to, following the rules, adhering to norms, putting forth necessary effort, persisting in the pursuit of appropriate goals, asking questions, and contributing to discussion.

In chapter 5, Riedl and Young argue the importance of narrative as an effective affective instructional strategy. Narrative is one of the fundamental modes used to understand the world around us. They compare the increasingly complex and difficult progression of skill-based activities in training/tutoring systems and games leading to skill mastery. They argue that it is not always enough to have a progression of skill-based activities. Games also use narratives to reinforce immersion within the game and motivate skill-based activities. As ITSs develop the ability to reason about and adapt storylines in response to learner needs, they will become more powerful instructional tools. However, developing narrative story-lines is currently a labor-intensive process. Today, automated story generation systems do not understand how the narratives they generate produce affective responses in learners. The major contribution of this chapter is that it identifies capability gaps and research challenges leading to automated, affect-sensitive narrative generation. More adaptive systems require more narrative content to support learner needs. Producing additional content is certainly more costly so automated narrative generation is significant not just to making ITSs more adaptable, but also in making authoring ITSs more cost effective.

Chapter 6, by Ritter, Sinatra, and Fancsali, notes a recent trend toward improving learning outcomes through a focus on "non-cognitive factors" such as learner motivation, beliefs about learning, learner interests, and metacognitive skills. They argue that personalization influences learner affect resulting in higher interest in the tutoring content and higher motivation to engage with that content. This leads to tighter focus and attention while feeling the enjoyment associated with achieving their goals. The authors recommend that GIFT could benefit from a wide variety of options in collecting personalization information from the learner rather than the traditional survey. The major contribution of this chapter is that it

identifies a variety of personalization techniques that enhance performance by engaging and motivating the learner.

In chapter 7, Arroyo, Muldner, Burleson, and Woolf examine adaptive interventions to address learner negative activating and deactivating emotions during instruction. The ability of the ITS to classify a learner's emotional state is a critical step toward adaptive instruction tailored to each learner's affective needs. This statement has been the focus of much research and aligns with the model portrayed by Sottilare's adaptive tutoring learning effect chain. However, little research exists on systematically examining the influence of affective interventions on learning outcomes (e.g., performance, knowledge and skill acquisition or retention), affect, and attitudes. In other words, how ITSs should to respond to learner emotions to provide optimal learning experiences in both the short and long term. The authors argue for three interventions to aid in managing learner emotions, which are the major contributions of this chapter: (1) target the learner's beliefs about the self and the task (value-oriented interventions), (2) target learner's self-regulation strategies to help them self-regulate their emotions and their learning process in effective ways (control-oriented interventions), and/or (3) manipulate the learning context (context-oriented interventions) to keep learners within Vygotsky's ZPD.

Finally, Chapter 8, by Ventura, Shute, and Small, describes the importance of assessing persistence in educational games to enhance learning. Their targeted educational game is Newton's Playground (NP), a two-dimensional, computer-based game aimed at helping learners understand qualitative physics. As part of the authors' validation process, they administered a performance-based measure of persistence consisting of impossible anagrams (jumbled letters that do not make a word) and impossible picture comparison tasks (two adjacent pictures where participants are told to detect difference between pictures when in fact no differences exist). The major contribution of this chapter is that it provides empirical research to define the positive relationship between persistence and learning in educational games where none existed before. An advantage of the methods described within this chapter is the simplicity of the data needed to detect learner persistence and the ability to apply these methods to a variety of educational settings. This bodes well for integration of persistence detection into generalized tutoring architectures like GIFT.

The Future of GIFT as an Instructional Manager

The contributors to this section of the book offer recommendations for developing instructional techniques, strategies, and tactics within GIFT. These recommendations address substantial challenges and opportunities that are envisioned to evolve over an extended period of time due to their complexity. The following enumerates recommended actions for consideration in the long-term view of GIFT as an instructional manager. Some of these recommended actions are already defined with known value (technology pull) and some are more speculative (technology push) in that their impact is difficult to predict at this time without additional empirical research.

1. *Improve Methods for Selecting Optimal Strategies for Learners*: Significant effort has been expended to develop affect detectors for both individual learners and teams of learners. A systematic analysis based on empirical studies should be conducted to evolve methods for selecting optimal instructional techniques, strategies, and tactics to manage affect, enhance engagement, and develop grit (perseverance, persistence, and resilience). Optimal instructional selection methods should be integrated within GIFT to enhance the performance of its existing engine for eMAP for strategy selection and enhance methods within the domain module for selecting tactics based on strategy selection.

2. *Reexamine and Integrate ITS Principles into GIFT*: The ITS design principles delineated by Anderson, Boyle, Farrell & Reiser (1987) and later elaborated by Corbett, Koedinger and Anderson

(1997) have been primarily focused on reacting to the cognitive states of the learner. We recommend effort be put forth to research and expand the methods for implementing these ITS design principles from both a cognitive and affective perspective, as noted in Chapter 2 of this text.

3. *Improve the Emotional Intelligence of ITSs*: Instructional strategy selection begins with accurate assessment of the learner's state. In the case of affective states (e.g., emotions), we recommend effort be put forth to conduct research and expand the emotional intelligence of ITSs to improve their capability to assess and optimally manage learner affect.

4. *Model Expert Human Tutors in ITSs*: Extensive studies have been conducted to identify and model expert human tutors. We recommend that similar effort be put forth to capture the desirable behaviors and traits of human tutors for use in computer-based ITSs, thereby enhancing ITS credibility and improving the trust between learners and ITSs.

5. *Balance Learner Capabilities and Tutoring Content*: Few ITSs have captured the principles of Vygotsky's ZPD where the tutor seeks to match the capabilities of the learner with the complexity and challenge level of the learning experience. We recommend research be conducted to expand the methods for implementing principles of ZPD within ITSs.

6. *Automate Narrative Generation*: As expectations for ITSs to be more adaptive grow, more and more narrative content to support learner needs will be required. We recommend research be conducted to accelerate the automation of narrative generation. We suggest this is not just significant to making ITSs more adaptable, but also to making authoring of ITSs more cost effective. It should be a goal to integrate narrative automation and narrative retrieval capabilities into GIFT.

7. *Use Non-Cognitive Factors to Enhance Engagement and Motivation*: Learner affect has been shown to be a moderator during the learning process. Likewise learner interests have been shown to influence engagement and motivation. Efforts should be continued to evaluate the relationship of non-cognitive factors and optimal instructional strategy selection within ITSs. Methods to easily integrate affect detectors, store and retrieve long-term traits and trends from learning record stores, and develop the ability to adapt narrative to include learner interests are all desirable capabilities for GIFT.

CHAPTER 2 – A Guide to Instructional Techniques, Strategies and Tactics to Manage Learner Affect, Engagement, and Grit

Robert A. Sottilare[1], Jeanine A. DeFalco[2] and Janice Connor[3]
[1]U.S. Army Research Laboratory, [2]Teachers College, Columbia University, [3]University of Central Florida

Introduction

This chapter reviews instructional methods (techniques, strategies, and tactics) in the literature used to promote positive affective states resulting in the acquisition of "robust" knowledge and skill development; mitigate negative affective states, which inhibit learning; enhance and maintain learner engagement to maximize opportunities for learning; and support the development of learner perseverance or grit. For purposes of our discussion, an *instructional technique* is a domain-independent and largely learner-independent method used in either human- or computer-based tutoring. In other words, the technique is implemented within an ITS as a method that has been shown to have positive effect on learning across training and educational domains, and across a variety of learners and learner states and traits. Instructional techniques are best practices developed over time and include, but are not limited to error-sensitive feedback, mastery learning, adaptive spacing and repetition, metacognitive prompting, and faded worked examples. Generalized instructional techniques are discussed later in this chapter.

An *instructional strategy* is a domain-independent plan or recommendation used by the tutor to guide the learner or adapt the level of challenge during tutoring. Instructional strategy selection may be informed by specific learner states (e.g., affect, performance) and/or learner traits (e.g., goal orientation). Instructional strategies may be developed *a priori* (based on learner information acquired prior to, or at the initiation of, instruction) or *in situ* (based on learner information acquired in real time during instruction).

Macro-adaptive strategies are *a priori* strategies that account for historical information about the learner including previous domains under training, achievements, and experience that aid in identifying the learner's competence in the current domain under training. Learner data to support macro-adaptive tutoring decisions may be acquired from Learning Record Stores (LRSs) or other online repositories. Micro-adaptive strategies are *in situ* strategies, which rely primarily upon real-time data streams, but may also use historical data in planning and instructional decision making.

An *instructional tactic* is a domain- and learner-dependent action taken by the ITS and may include the presentation of hints, prompts, questions, assertions, questions, and other tutor-initiated behavior and responses. Instructional tactics within GIFT are actions to be taken by the ITS (e.g., provide information; offer specific feedback; prompt for learner reflection on a specific concept). Instructional strategies narrow the available options for instructional tactics. For example, an instructional strategy may be a plan to ask the learner a question to test their knowledge about the domain under training. If the current instructional context is that the learner is being tutored about marksmanship principles, then the ITS would implement a tactic by retrieving a specific question about marksmanship from the available libraries and then present that question to the learner via text or voice.

As previously noted in the prologue of this volume (Nye, Sottilare, Ragusa & Hoffman, 2014), an *instructional strategy* develops plans and makes decisions that are intended to improve the learner's knowledge and skill within the domain tutored by the ITS. Instructional strategies are intended to implemented in near real time, but their data, upon which the ITS makes instructional decisions, may also be

historical. Instructional strategies may be tied to a near-term objective (e.g., acquire knowledge) or a long-term objective (e.g., promote deep learning to enhance retention). The effectiveness of instructional decisions in promoting truly adaptive tutoring is tied directly to the ITS's knowledge and perception of the learner's traits, behaviors (actions, demonstration of progress), and physiology (response the learning environment).

Generalized Instructional Techniques

Instructional techniques have been derived over time as best practices due to their successful application. The instructional techniques presented below are for reference and provide a sampling of effective techniques based on the literature. They are not intended to be a complete set of all available techniques implemented, but do include the most used techniques.

Durlach and Ray (2011) conducted an extensive review of over 200 studies that explored the effectiveness of adaptive versus non-adaptive instructional techniques in computer-based tutoring systems. They identified the most adaptive form of instruction as "model-based adaptation" in which the computer adapts to both the local input from the student but also from other information about the learner such as physiological data, prior experience, and preferences. Their review of stringent criteria resulted in 17 studies that show that adaptive versus non-adaptive instruction is superior for learning (Durlach, 2011). Other studies have shown adaptive computer-based systems, such as the Lisp Tutor (Anderson, Corbett, Koedinger & Pelletier, 1995), AutoTutor (Graesser et al., 2004), and the Andes Physics Tutor (VanLehn et al., 2005) increase learning with effect sizes ranging from 0.80 to 1.05 σ.

Error-Sensitive Feedback

Error sensitive feedback is a technique where an intervention is triggered when the learner commits errors that are either individually or cumulatively significantly divergent from the ideal as defined in the expert model of the domain within the ITS. Implementation of error-sensitive feedback with ITSs poses the significant challenge of providing timely feedback while maintaining flow/engagement. High frequency feedback resulting from several errors may result in negative learner affect (e.g., frustration).

Error-sensitive feedback may be given when a learner incorrectly answers a question or seems unsure of a correct answer, as determined by amount of time to answer question (latency) or repeated requests for help. Feedback is specific to the answer selected, discusses common misconceptions that may have led to the incorrect answer, and steers the student to absorb the information and self-reflect on their answer and their reason for selecting it. Although feedback has been shown to be effective for learning, the difficulty in computer-based tutoring is determining at what frequency to deliver the feedback and also determining why the learner might have erred.

According to Durlach and Ray (2011), error-sensitive feedback might be helpful if the student erred because they simply forgot the material; it might not be helpful if the learner does not comprehend the material – no reminder or review will ultimately help lost learners find their way and will ultimately lead to frustration (p. 24).

Shute (2007) discussed differences in philosophy in immediate versus delayed feedback. There are two schools of thought – one says to provide feedback on the error as soon as it occurs so the error does not become retained; proponents of delayed feedback advocate waiting and reinforcing corrective behavior – the original error will be quickly forgotten once understanding of mistake takes place. Both approaches have been shown to be effective (Shute, 2007). Immediate feedback may be more helpful on difficult tasks; delayed on easier tasks where easy and difficult are defined by the learner's domain competency.

Overall, however, results on delayed versus immediate feedback are situational. In addition, long, extensive feedback may lead to learner frustration resulting in negative effects on learning and decreased recall (Shute, 2007).

Mastery Learning

Mastery learning is a technique where the tutor insures the learner has "mastered" (can recall and apply) prerequisite knowledge before allowing the learner to move on to the next lesson/concept: "Mastery does not imply perfection, but satisfactory performance" (Murray & Arroyo, 2002). In this context, satisfactory performance is generally defined as the minimum standard to pass. In this way, mastery learning may contribute to higher self-esteem based on achievement or may contribute to frustration if the learner is unable to grasp requisite concepts and does not move forward in the curriculum.

Adaptive Spacing and Repetition

Adaptive spacing and repetition is a technique where the learner more easily recalls knowledge items/objects when these knowledge artifacts are exposed to the learner repeatedly over a long time span rather than repeatedly studied during a short time span (Dempster, 1988). This prolonged exposure promotes deeper learning and extends the spacing between instances of refresher training.

Metacognitive Prompting

Metacognitive prompting is a technique where the tutor encourages the learner to self-reflect and evaluate, self-explain, and self-correct rather than provide the answer directly. The Cognitive Transformation Theory (CTT; Klein & Baxter, 2006) asserts that problem solving on the part of learner involves the recognition of flaws in their existing mental models and restructuring of those models by shedding flawed elements of those mental models for less-flawed models through reflection and discovery. Sottilare & Goldberg (2012) note the need for the tutor to support processes, which allow the learner to construct and restructure their own mental model in order to promote transfer and, in some cases, accelerate learning.

Fading Worked Examples

Fading worked examples is technique where the tutor provides "a step-by-step demonstration of how to perform a task or how to solve a problem" (Clark, Nguyen & Sweller, 2006, p. 190), from which parts have been deliberately removed or faded (Atkinson, Renkl & Merrill, 2003). This technique challenges learners to recall and reconstruct their mental model of the task or concept in much the same way as metacognitive prompting, but provides additional context needed for novices to recall the missing elements. This technique is especially applicable to tutoring where problems are presented to the learner to solve.

Subsequent sections of this chapter address the three primary themes within the literature (affect, engagement, and grit) and their relationship to instructional management principles and learning outcomes, but first we identify and discuss the criteria by which we will evaluate the effectiveness of reviewed instructional strategies and tactics.

Assessment Criteria for Adaptive Instructional Techniques, Strategies, and Tactics

In assessing the value of adaptive instructional strategies, we examined their impact in terms of their influence on the desirable outcomes noted below. First and foremost is learning. The next three (modera-

tion of affect; influence on motivation and engagement; and development of desirable traits) all influence learning. The last criterion (ease of authoring and reusability) addresses practical considerations and influences the cost and time to produce/maintain ITSs.

Learning Effect

Since the goal of ITSs is to promote learning in one-to-one and one-to-many instructional situations, we considered the degree to which instructional strategies and tactics affect learning outcomes as the most important assessment criterion. For purposes of this discussion, we consider *learning* as a relatively permanent change in, or acquisition of, knowledge, skills, understanding, or behavior.

Effect size was chosen as a method to compare disparate instructional strategies in terms of their impact on the following learning outcomes: knowledge and skill acquisition, acceleration of learning, performance, and retention. Performance accounts for the transfer of learning from one context to another. Per Byrnes (1996), instructional practices considered critical to learning and performance include providing multiple contexts for the original learning; representing problems at higher levels of abstraction; overlapping the original domain of learning and the new one to a high degree; and implementing dynamic processes that require learners to actively choose and evaluate strategies, consider resources, and receive feedback.

Moderation of Affect

A second criterion considered in our review accounts for the influence of instructional strategies and tactics is managing affect (personality, mood, and emotions). According to Gebhard (2005), personality, mood, and emotions vary in duration, influence, and cause. The influence of affect on learning may be positive or negative. The relationship between affect and learning is well documented in terms of affect's influence on accepting new knowledge (Linnenbrink & Pintrich, 2002), creativity in problem solving (Isen, Daubman & Nowicki, 1987; Isen, 2000; Isen, 2003; Isen, 2004; Isen & Erez, 2006), enhancing interaction (Norman, 2002), and recall (Gold & van Buskirk, 1975; Bower, 1981; Bower & Forgas, 2000).

Influence on Motivation and Engagement

The third criterion in our review addresses the design of ITSs to influence motivation and engagement. The relationship between motivation, engagement, and learning are well documented, (Corno & Mandinach, 1983; Kearsley & Shniederman, 1998; Blumenfeld, Kempler & Krajcik, 2006; Pugh, Linnebrink-Garcia, Koskey, Stewart & Manzey, 2010). Moderation of affect may be used to enhance motivation, (Erez & Isen, 2002), and artificially intelligent agents have been used within ITSs to influence learner interest and motivation (Rosenberg-Kima, Plant, Baylor & Doerr, 2007).

Development of Desirable Traits

As a consequence of effective ITS design, desirable traits may result in enhancement of non-traditional learning outcomes such as learner creativity, adaptability, and perseverance (also known as grit). Development of these traits may aid the learner in transferring knowledge and applying skills in other domains (Byrnes, 1996).

Ease of Authoring and Reusability

Instructional strategies and tactics may be effective, but the ability to easily author or implement strategies in new ITSs directly determines their practicality and ubiquitousness. Ease of authoring may also be directly tied to their reusability, generalizability, or domain-independence of strategies across different educational and training domains. The authoring process may be complicated by the level of definition (well-defined or ill-defined) of the domain to be tutored. Modular approaches to the development and application of instructional strategies within ITSs that promote standards and best practices are desirable.

To better understand the effectiveness of strategies and tactics on affect, engagement, and grit, the next section of this chapter reviews various ITSs strategies and tactics selection methods. The goal of this review is to compare/contrast selection criteria and consider the strengths and weaknesses of each selection method to help researchers and instructional system designers identify best practices for incorporation in ITS designs and architectures like GIFT.

Literature Review Themes

This section is dedicated to a review of the literature in three areas: instructional strategies/tactics to manage learner affect and enhance learning; instructional strategies/tactics to enhance engagement, and thereby, learning; and instructional strategies/tactics to enhance the desirable trait of grit or perseverance. Since subsequent chapters will address specific strategies in these areas, our intention is to provide an overview of the literature relative to selected constructs or models. This is not intended to be a comprehensive assessment of the literature, but the examples chosen should provide general knowledge of effective instructional strategies and tactics to a point sufficient for the reader to construct an initial mental model.

Exploring Methods to Manage Learner Affect

What follows is a review of approaches to moderating learner affect to optimize engagement and motivation, and thereby, learning per our stated criteria. However, before we delve into methods used to manage or influence affect, it is necessary to define "affect" and its associated states. According to the American Psychological Association (2007), *affect* is the subjective experience of feeling. As mentioned previously, Gebhard (2005) generalized three affective states within his model A Layered Model of Affect (ALMA): personality preferences, mood, and emotions, which differ in their duration, influence, and cause. The level of affect is defined in terms of its valance and arousal. Valence is a subjective positive-negative evaluation of experience, and arousal is activation to action or physiological readiness to take action. Arousal ranges from "excited" indicated by high physical activity and mental alertness to "calmness" indicated by low physical activity and mental sluggishness (Klesen, 2002).

When we discuss instructional strategies for ITSs, it is essential that we identify desirable salient characteristics and capabilities of ITSs to manage the learning process of the user. A critical capability of a tutor is its ability to manage affect to motivate the learner and improve the learning process (Hernandez, Noguez, Sucar & Arroyo-Figueroa, 2006). Picard (2006) identified essential capabilities for a tutoring system to manage affect: 1) accurately recognize the student's affective state; 2) respond appropriately to the student's affective state; and 3) understand when and how to appropriately express emotion to build trust and motivate the student. Additional desirable capabilities for ITSs relative to managing affect (e.g., avoiding frustration) include modifying the presentation of information so that learning proceeds efficiently (Johnson & Taatgen, 2005); identifying and responding to the learner's affective cues in a timely fashion (Alexander, Sarrafzadeh & Fan, 2003); and delivering content in a way that adapts to each

learner's particular preferences (Rodrigues, Novais & Santos, 2005) and their domain competency level (Sessink, Beeftink, Tramper & Hartog, 2007).

Since subsequent chapters within this section of our book review specific strategies about *responding to affect* (Chapter 3; D'Mello, Blanchard & Baker), addressing *disengaged behaviors and affect* (Chapter 4; DeFalco, Baker & D'Mello), and using *narrative as an affective instructional strategy* (Chapter 5; Riedl & Young), this literature review theme is focused on very specific areas to promote complementary information and discussion: *developing emotional intelligence* in adaptive tutoring systems; modeling the response of *expert human tutors* to moderate (promote positive results and reduce negative consequences) learner affect; and using the *zone of proximal development* (Vygotsky, 1978) as a model to manage affect.

Review: Developing Emotional Intelligence in Adaptive Tutoring Systems

"The extent to which emotional upsets can interfere with mental life is no news to teachers. Students who are anxious, angry, or depressed don't learn; people who are caught in these states do not take in information efficiently or deal with it well" (Goleman, 1995). Understanding and managing emotions (our own and others) is critical to our academic success and, like their human counterparts, it is important for ITSs to be able to detect, identify, and manage each learner's emotions to optimize learning. To a large extent, this includes maintaining the positive affect of the learner, portraying positive affect, and avoiding long-term negative affect (e.g., confusion, frustration, anger), which detracts from learning. D'Mello and Graesser (2012) note confusion as a key indicator of cognitive disequilibrium, which occurs when a learner reaches an impasse. Learners must then exert significant effort to solve the problem in order to resolve the impasse and restore equilibrium (flow/engagement) within the learning process. This temporary impasse caused by confusion can, in fact, be good for learning, but may have the opposite effect if confusion is allowed to persist, since equilibrium is not restored.

Picard (2006) notes the inadequate adaptability of ITSs to the needs of learners. This is due in large part to the inability of ITSs to accurately and unobtrusively classify emotions during one-to-one tutoring. While significant progress has been made since 2006, more accurate (>95%) real-time predictive models of affect are needed to support the adaptive tutoring learning effect chain first described by Sottilare (2012). Managing emotions starts with recognizing emotions. Even highly accurate learner emotional state classifiers (e.g., 98%) may introduce significant error (~27%) if the dependent strategy and tactics selection classifiers in the chain are only 80% accurate (Sottilare, Ragusa, Hoffman & Goldberg, 2013). Due to interdependencies between learner state classification, instructional strategy recommendations, and instructional tactics selection, each of the classifiers in the chain must be highly accurate and in real time to manage learner emotions and support truly adaptive tutoring tailored for each individual learner's needs. The following publications note instructional strategies and tactics to enhance the emotional intelligence of tutors or the perceptions of the learner while interacting with computer-based systems (e.g., tutors, dialogue-based webpages).

Andre, Rehm, Minker, and Bühler (2004) investigated methods to improve the user's perception of the interaction with a dialogue-based system by integrating social models of politeness and cognitive models of emotions. The resulting model influenced the subjective perception of the interaction between the user and the system to a large extent based on the duration of the relationship between the user and the system. Four primary strategies were used based largely on the assessment of the user's affective state (valence and arousal): direct, approval-oriented, autonomy-oriented, or off the record. The idea of using conversational agents (also known as virtual characters or chatbots) to promote bonding between machine and user is now widely used in e-marketing to gain the confidence of the potential buyers.

In interactions with robots, Hoffmann and Krämer (2011) also noted learner affect and engagement may have been related to attributes of the robot including, but not limited to robot size, realism, shared

physical space, physical presence, and perceived social presence. Dialogue-based tutors (e.g., AutoTutor) also seek to enhance the relationship between the learner and the tutor (Lane & Johnson, 2008) in an effort to enhance engagement, and thereby, increase opportunities for learning. Finally, Embodied-Perceptive Tutors for Empathy-Based Learning (EMOTE) is a collaborative project that aims to develop artificial tutors capable of emotionally engaging with learners (Corrigan, Peters & Castellano, 2013). Implementing emotional intelligence may be considered moderate to difficult depending on system goals.

Review: Modeling the Response of Expert Human Tutors to Learner Affect

Lepper, Drake, and O'Donnell-Johnson (1997) identified the characteristics of an ideal human tutor in their intelligent, nurturant, Socratic, progressive, indirect, reflective, and encouraging (INSPIRE) model. The characteristics in this model were further explored in Lepper and Woolverton's (2002) study of highly effective tutors with the goal to differentiate best tutoring practices. It was found that highly skilled tutors manage two primary processes: engagement and motivation. Some of the findings regarding the tutor's response to learner affect relate primarily to the following tutor characteristics: nurturant, indirect, and encouraging.

Nurturant tutors established rapport with the learner early, demonstrated empathy when the learner experienced significant difficulties, and vocalized confidence in the learner's ability to succeed. These behaviors may be difficult to replicate in computer-based tutors without significant knowledge of the learners: their interests, their past achievements, their capabilities, and real-time assessment of their performance and emotional states. While the type of feedback/scaffolding is important to success, the frequency of interaction may also be important. The learner may perceive too frequent praising by the tutor as insincere or annoying (Person, Kreuz, Zwaan & Graesser, 1995). Indirect tutors avoid overt criticism and explicit praise. They may imply the existence/location of an error and then prompt the learner to review their own mental model, reflect, and find their own mistakes. Finally, encouraging tutors manage the confidence and curiosity of the learner along with the challenge level of the learning experience to maintain motivation, and thereby, influence positive affect and perseverance.

Additional insight in how human tutoring principles might be implemented in computer-based tutors might best begin with Anderson, Boyle, Farrell, and Reiser's (1987) principles for ITS design as later elaborated by Corbett, Koedinger, and Anderson (1997). While these principles have a cognitive focus, we discuss how implementation (or lack of implementation) of each principle might influence learner affect (including motivation):

Principle 0: An intelligent tutor system should enable the student to work to the successful conclusion of problem solving. While Corbett et al. (1997) demonstrated the importance of enabling the student to work to a successful conclusion from a cognitive point of view, there are affective and motivational reasons to insure that learners have the opportunity to reach a successful conclusion. If the problems presented are too difficult to finish, the lack of achievement may cause learners to withdraw due to frustration. It is important for the tutor to engage the learner and keep them engaged. By keeping the learner engaged, the greatest opportunity for learning exists. If the tutor fails to keep the learner engaged, there is virtually no opportunity for learning. This idea of matching problem difficulty to learner competence is discussed in more detail in the next subsection on Vygotsky's ZPD.

Principle 1: Represent student competence as a production set. By representing the learner's procedural knowledge in a structured way, it is possible to match learner competence to appropriately challenging problem sets to keep the learner engaged and reduce affective byproducts such as boredom or frustration. Understanding and modeling learner competence may be easier in very well-defined domains where goals are very specific and there are generally just a few or maybe only one path to achievement. It may prove

to be more difficult in ill-defined domains where information needed to assess competency is not easily agreed upon or where there are multiple paths to success.

Principle 2: Communicate the goal structure underlying the problem solving. By dividing the problem into smaller subgoals and tying domain knowledge to these sub-goals, the tutor can guide learning rather than directly provide answers. When learners ask for help, the first level of help for each problem is a reminder of the current state of the problem and the desired sub-goal state. Subsequent help messages provide advice about how to accomplish the goal/sub-goal. This encourages the learner to apply their newly acquired domain knowledge toward the goal while maintaining interaction with the tutor. This communication may be critical in developing rapport with the tutor as the learner attains these goals. While the implementation of this principle should be considered easy for well-defined domains (e.g., mathematics, physics), it may prove to be much more difficult to implement for ill-defined domains where there are multiple paths to successful outcomes.

Principle 3: Provide instruction in the problem-solving context. This principle coincides with Merrill's Component Display Theory (CDT; Merrill, Reiser, Ranney & Trafton, 1992), which examines methods for effective tutoring techniques. Merrill's CDT discusses a process for presenting rules, providing examples, testing for recall of knowledge, and applying knowledge in guided practice. CDT's phase of guided practice aligns with this principle's problem-solving context in that learners are presented with a situation in which they must use their knowledge to reach a successful outcome (e.g., complete the problem). Problem-solving contexts are a vehicle for deeper learning rather than rote memorization of facts in that they allow the learner to apply knowledge, and build and rebuild mental models in an effort to generalize their learning. This may be important in maintaining engagement and motivation since it demonstrates real-world application of knowledge while avoiding the learner's question "why did I learn this?" This principle should be considered easy to implement and domain-independent, making it reusable across multiple training domains.

Principle 4: Promote an abstract understanding of the problem-solving knowledge. The goal is to "encode problem-solving states, actions, and consequences in the most general form consistent with the problem-solving goal in order to foster transfer across contexts," (Corbett et al., 1997). This mental model allows learners to apply knowledge learned in one context to another context while it expands the potential usefulness of learning, and thereby, the motivation of the learner (Pea, 1988). This principle should be considered the most abstract and hardest to implement within a tutoring system depending upon the states/traits tracked in the learner model. Differences in how each individual constructs/deconstructs/reconstructs a mental model may contribute to the difficulty in implementing this principle in more complex and ill-defined training domains.

Principle 5: Minimize working memory load. To manage working memory load, "human tutors do not interrupt students (and their current working memory state) to point out relatively minor errors that have little consequence for the student's overall goal structure" (Corbett et al. 1997). Interruptions of any kind can impact the learning process negatively by increasing working memory load. This principle can also be applied to the acquisition of learner data noted in the adaptive tutoring learning effect chain, (Sottilare, Ragusa, Hoffman & Goldberg, 2013), when considering sensors for behavioral and physiological data collection. To keep working memory load focused on germane data, the sensors employed to acquire learner data should be unobtrusive. Sensors that are uncomfortable to wear or restrict movement should be avoided as they may add to working memory load and learner frustration due to an inability to focus attention on the problem at hand. At the very least, this is a diversion of the learner's attentional focus. This principle should be considered easy to implement and domain-independent, making it reusable across multiple training domains.

Principle 6: Provide immediate feedback on errors. Corbett et al. (1997) admit that this principle is controversial due to its potential impact on working memory load. So in order to bring it back into balance, we recommend changing it to read: "*provide immediate feedback on **critical errors**.*" We define critical errors to be errors leading to catastrophic outcomes (e.g., large-scale loss of time or significantly reduced learning). By only providing feedback on critical errors, this principle remains in balance with principle 5 by minimizing the effect on working memory load and reducing potential for learner frustration. Additional research is needed to identify thresholds for defining critical errors. This principle should be considered moderately difficult to implement and may not be domain-independent. Incorrect implementation (too little or too much feedback) might be especially frustrating for novice domain learners. The "too little feedback" implementation scenario could indirectly contribute to identification of learner grit or perseverance.

Principle 7: Adjust the grain size of instruction with learning. Based on principle 5, minimizing working memory load, this principle recommends adjusting goal decompositions differently based on the competency/experience of the learner. This principle suggests that more granular (smaller) decompositions are needed for beginners, but more experienced learners should be able to focus on higher level goals and decompose them on their own, thereby reducing working memory load. This principle parallels Vygotsky's ZPD (1978), which advocates alignment of problem challenge levels with the learner's competency/experience, and is likely to produce a tutor that induces less learner frustration. This principle should be considered easy to implement and domain-independent, making it reusable across multiple training domains.

Principle 8: Facilitate successive approximations to the target skill. This principle calls for a reduction in the amount of scaffolding or support as the learner becomes more proficient over time. In other words, more direction and feedback are required to keep the novice on track, and less is needed for more experienced domain learners. This does two things for the learner: 1) reduces the frequency of interruptions during the learning process; and 2) signals that the tutor has growing confidence in the learner's problem-solve ability. These tutor behaviors and learner perceptions coalesce to form trusting relationships between learners and tutoring systems over time. Inversely, the failure to reduce scaffolding over time might diminish trust and willingness of the learner to take risks, thereby stunting the development of creativity as it relates to problem solving. This principle should be considered easy to implement and domain-independent, making it reusable/transferrable across multiple training domains.

Review: Using the Zone of Proximal Development to Manage Affect

The ZPD (Vygotsky, 1978) purposefully matches learner competence and the challenge level of the problem or tutoring experience. As in the adaptive tutoring learning effect chain (Sottilare, 2012; Fletcher & Sottilare, 2013; Sottilare, Ragusa, Hoffman & Goldberg, 2013), the key to applying optimal strategies is the ability to accurately classify the learner's affective state. In Vygotsky's model, affect as a temporary disequilibrium (e.g., confusion) is an indicator of a mismatch between the competency of the learner and the problem or challenge presented by the tutor. The job of the ITS is to adapt the instruction by responding to the learner's state. Murray and Arroyo (2002) observed that ITSs can adapt at three levels: sequencing content, providing opportunities for practice, and giving feedback. Perhaps there may be other levels of adaptation (e.g., provide opportunities for reflection), but Murray and Arroyo captured the essence of tutor options in response to classification of affective states as *specific ZPD* (*SZPD*). SZPD is composed of three factors: H – the goal number of hints allowed in each problem set; DH – the allowed variation in H to consider the current state to be within the ZPD; and P – the minimum number of problems the learner is guaranteed to attempt. The SZPD factors, then, are properties of the instructional strategy and can be used for post-hoc analysis or dynamic adaptation of strategies based upon the relative effectiveness of different hinting styles.

Summation for Managing Learner Affect

In this section, we have specifically segregated the *managing learner affect* problem space into two distinct processes: detecting affect and selecting appropriate strategies based on the learner's affective state. With respect to detecting affect, we highly recommend reading Calvo and D'Mello's (2010) interdisciplinary review of affect detection.

For the task of selecting optimal strategies based on the learner's affective state, we examined three approaches: developing emotional intelligence in adaptive tutoring systems; modeling the response of expert human tutors to learner affect; and finally, using the ZPD to manage affect.

Developing the emotional intelligence of ITSs is largely dependent upon recognizing emotions and, perhaps more importantly, any disequilibrium over extended periods of time that might prove detrimental to learning. The use of dialogue-based systems seems to be one way to foster human confidence and trust in computer-based tutors. Additional research is needed to fully understand and model the attributes of the dialogue-based tutors, which have the most influence in promoting and maintaining positive affect.

Modeling the response of expert human tutors to affect for use in computer-based tutors is a complex and multi-dimensional task. We reviewed Lepper's INSPIRE model and reexamined Anderson's principles for ITS design from the perspective of how each principle might influence or manage affect.

Finally, we investigated methods for modeling the ZPD within ITSs and found Murray and Arroyo (2002) had operationalized definitions and assessments for implementing the ZPD within ITS. Specific measures of hint requests were used to assess flow.

Adaptive Instructional Strategies to Enhance Learner Engagement

The digital revolution is changing the face of education, and yet the fundamental constructs of how people learn remain the same. Further, some would argue that the purpose of education itself is changing from teaching facts and figures to developing the mind of the learner. Indeed, educational researchers such as Sternberg note that instruction should not be geared solely toward imparting a knowledge base, but developing practical, creative, reflective, and analytical skills (Sternberg, 1998). The education of the learner, then, necessitates pedagogical designs that are both an art and a science, as Dewey noted almost 100 years ago. It is the interplay between choice (art) and methodology (science) of instruction that ultimately promotes the phenomenon known as engagement necessary to facilitate learning.

As such, before reviewing best instructional practices in the classroom as a model for technology-based learning platforms, addressing the why (philosophy), the how (science), and the meeting place of the two are necessary first steps in this review. Following that analysis, this theme reviews current instructional practices that support the engagement of students, as well as provides some specific strategies that can be parlayed into an online learning platform.

The Why: Philosophy of Education

Developing the mind of the learner has been the battle call of progressive educators for over a century now, most notably as seen in the voluminous works of John Dewey. For Dewey, teaching and learning were not only an obvious necessity for living but education was the instrument through which to promote a democratic society (Dewey, 1944).

Democracy can only be sustained, asserts Dewey, in a society where the shared interests of the group are maintained through an education that promotes both the personal interests of its members as well as develops the habits of mind that secure social change without disorder (Dewey, 1944). Developing the habits of the mind for intelligent living where all people are problem solvers, then, was the chief aim of education for Dewey (Eldridge, 1998). This aim seeks to make education a mode of practice and not just theory (Dewey, 1929).

The concept of the practice of education connotes an effort and engagement of the learner, and is fundamental to the idea of Dewey's notion of experiential learning. Dewey's views, in combination with the theories of psychologists such as Piaget and Vygotsky, have contributed to the constructivist movement in education that is widely accepted as the dominant school of thought shaping curriculum design of the classrooms today (Hickman, Neubert & Reich, 2009).

Constructivist Perspective

Constructivism is more than a theory of learning. It is a philosophical approach to investigating the structure, scope, and nature of knowledge (Pritchard & Woollard, 2010). Social constructivist theory emphasizes that human learning and behavior occur in social environments. This theory is based on the central notion that as learners we construct our own knowledge about the world around us based on our experiences and the interaction of others (Schunk & Mullen, 2012; Pritchard & Woollard, 2010). The construction of this knowledge begins when we connect our past knowledge with new and current knowledge, transforming that experience into new, personal knowledge and understanding (Pritchard & Woollard, 2010).

The implications of constructivism on education praxis, then, includes scaffolding and guided learning, identification of learner's strengths and or intelligences, individual learning plans, problem-based learning, diagnosis of individual learning styles, and incorporating learners' views (Jordan, Orison & Stack, 2008). The instructional implications also includes using raw data and primary sources, providing physical, interactive, and manipulative materials, creating opportunities for exploratory classroom discussion, and engaging pupils in experiences that might challenge previously held beliefs or understanding of phenomenon (Jordan et al., 2008).

Peer learning and the use of different teaching styles to aid the learner in interpreting the world are also core to the constructivist approach (Jordan et al., 2008). For the purposes of verifying best instructional practices, however, the constructivist pedagogy benefits from an examination through the lens of current cognitive science research that begins with how the brain encodes information, creates internal representations, and retrieves this information from memory (Strack & Forster, 2009).

The How: Cognitive Science and Education

From the fields of neurology and cognitive psychology, the cognitivist approach to instructional design is rooted in research that has identified five basic processes involved in cognition: sensation, perception, attention, encoding, and memory (Jensen, 2005). Sensation includes visual, audio, and haptic; perception includes pattern recognition and object recognition; attention includes how we focus limited mental resources at a given time while ignoring others; encoding refers to organizing information into mental representations or schemas; and memory, both working and long-term, factor into our ability to retain and recall information (Jensen, 2005).

In contrast to the constructivist perspective where learning is driven by the social experience of the learner, cognitivists maintain that learning involves developing effective ways of building schemata and processing information. This constitutes a process whereby the teacher is in control of learning and

meaning, it is up to the teacher to design materials that stimulate the learner's cognitive processes through which the learner is encouraged to make mental connections (Howard-Jones, 2010; Jensen, 2005).

Neuroscience has informed educational research that learners have a limited amount of working memory or capacity to hold information in attention when they we are processing it (Howard-Jones, 2010). External representations can help offload some initial working memory demands when engaged with new problems, a notion well articulated by the Cognitive Load Theory (Chandler & Sweller, 1991), which has formed an important basis for much instructional design.

Brain imaging explains that when observing others carrying out actions, some of the same cortical regions activate as if we were carrying out the actions ourselves (Rizzolatti & Craighero, 2004). This so-called mirror neuron system also activates when we hear of human actions being performed, which suggests how the narrative constructs and visualization can support learning (Howard-Jones, 2010). Neurobiological evidence also has illuminated how learning, attention, decision making, and social functioning are subsumed within, and affected by, emotional processes (Immordino-Yang & Damasio, 2007). For example, research on stress hormones reveal how stress facilitates memory when stress hormones are present at the time of learning, but have opposite effects when they are present before, or for an extended time after, the learning event (de Quervain et al., 2000).

Another interesting discovery in cognitive science research is the notion that creativity requires switching between two very different types of mental processes: generative and analytical thinking, each requiring a different attentional state. Analytical thinking is used to assess a problem or evaluate a potential solution and requires focused attention. Generative thinking, on the other hand, is needed to produce ideas and potential solutions, but needs more diffuse attention (Kounios et al., 2008). However, creative ability is not merely rooted in individual differences, but can be influenced in and by the instructional strategies of teachers in the classroom (Howard-Jones, Blakemore, Samuel, Summers & Claxton, 2005).

In terms of content retention and recall, research in cognitive psychology illuminates how emotion, sense, and meaning factor into enhancing memory. Barkley (2010) notes how information is more likely to be stored permanently if a learner makes an emotional connection to that information. Further, how well information makes sense to a learner will affect retention. The principle here being if there is a reason for the brain to remember information beyond just passing a test, a learner's ability to store and recall information will be enhanced (Barkley, 2010). Understanding how the brain receives, stores, and retrieves information, then, is a key element that must be considered when evaluating instructional designs of the classroom.

The Meeting Place: Engagement, Orientation, and Attention

The liminal space between constructivist epistemology and brain research is the phenomenon of engagement. Engagement in the classroom has become almost an unwieldy topic of research in education. To date, the examination of what constitutes engagement has been broken down into a number of elements that all seem to contribute to deep learning and transfer of knowledge. One of the seminal reviews of academic engagement is that of Fredricks, Blumenfeld, and Paris (2004) that define a three-part construct of school engagement that includes behavioral, emotional, and cognitive properties. Behavior engagement includes the participation in activities, effort, persistence, and positive conduct. Emotional engagement constitutes the positive and negative affective reactions such as frustration, boredom, and interest. Cognitive engagement covers the willingness of the learner to put forth the mental effort necessary to comprehend content and complete tasks across different learning domains.

Interestingly, Fredricks, Blumenfeld, and Paris (2004) summarize their review of engagement by suggesting that these three separate properties have not been studied in combination to each other nor have the

patterns or relationships between them. The authors pointedly suggest that engagement may in fact be a multi-dimensional construct influenced by social contexts and individual factors, a complex system of systems, both internal to the learner and in the environment (Fredricks, Blumenfeld & Paris., 2004). Indeed, this notion of engagement in education as a multi-dimensional construct calls to mind Bandura's social cognitive theory (1977, 1986, 1997, 2001) that speaks of the *triadic reciprocity* between personal factors, behaviors, and social factors, or what could be termed a phenomenology of engagement.

Over the last 30 years, researchers have focused on examining teacher-student interactions, the social conditions of a classroom, and the cognitive development of the learner (Reyes, Brackett, Rivers White & Salovey, 2012). The science of determining the how to promote engagement and learning helps direct our attention in both observation of and reflection on conditions and relationships that might otherwise go unnoticed. However, as Dewey pointed out, the end game of educational science lays in the minds of those engaged in directing educational activities through whom educational functions should become more intelligent (Dewey, 1929). It is to this end, then, that the examination of best practices in educational design has its great importance. For engagement does not rest merely in the preexisting conditions of a learner, or in content in isolation, or in the best of intentions of a teacher. Rather, engagement occurs when the learner interacts with content as structured by the instructor or teacher. One could argue that it is the way in which the learning experience is structured, or designed that is not only core to the learning process but it is through this structured experience that the triadic reciprocity, or phenomenon, of engagement occurs.

Constructivists argue that instructional design depends on the "creative genius of the teacher (the art and science of teaching); complex tools for instructional excellence (instructional methods); and expansive systems of interconnectivity to frame these learning experiences (curricular frameworks)" (Fogarty, 1999, p. 76). Cognitive scientists have known for some time that presenting material in both symbolic or pictorial form and literary text form enhances memory, and more recent evidence shows that multimodal stimulus produces additional brain activity beyond that experienced by each mode in isolation (Howard-Jones, 2010). Both epistemological views seem to support the notion that how material is presented not only stimulates additional brain activity, but it also serves to orient the attention of observer/student to best promote engagement in the learning process.

Instructional designs can be thought of, then, as an orienting process by the instructor for the learner. Orienting is the process of moving attention to a location, spatially or temporally orienting, and orienting attention possibly to particular stimulus, which co-occur in the same spatial location at the same time (Yiend, 2010). Orienting implies that stimuli or signals at a location and time become amplified, triggering the detection of the observer/learner toward a possibly significant event (Yiend, 2010). Ainley (2012) notes that a newly triggered situational interest involves arousal of affect and focused attention toward the object triggering interest, which in a novel situation begins a new mental schema. Further, Friedman, Fishback, Forster & Worth (2003) note "broad or narrow perceptual attention primes broad or narrow conceptual attention" (pp. 278-279).

As such, through a greater understanding of the importance of orientation and focus in instructional delivery, the superiority of the temporal tutoring model of teaching becomes more self-evident. As Noonan (2013) notes: "The most successful form of teaching involves the tutorial because of the obvious advantages, including selecting appropriate content and goals for the predicted stage of learner development and then modifying methods 'on the fly' based on the learner's response" (Noonan, 2013, p. 3). Being able to direct and focus a learner's attention on the gaps they experience in content or conceptual mastery is a key element in temporal tutoring. The process of directing the learner's focus to bridge these gaps and omissions further includes the selection and adaption of content and delivery as necessary corollaries of this tutoring model. As such, it stands to reason that a closer examination of instructional

best practices in temporal instructional contexts is central to the analysis of developing technology based learning platforms that promote engagement and learning.

Review: Best Practices of Instructional Designs

In reviewing the literature for best instructional practices in the classroom, there are generally three categories of sources. Firstly, there are individual papers in journals that examine a singular instructional design, i.e., activating haptic channels to promote learning (e.g., Chan & Black, 2006); self-explanations that promote cognitive change (e.g., Siegler & Chen, 2008); dialogic argumentation that develop adolescent thinking (e.g., Kuhn & Crowell, 2011); mental models that improve learning (e.g., Bucciarelli, 2004); and the process of inventing using contrasting cases (e.g., Schwartz, Chase, Oppezzo & Chin, 2011). Secondarily, there are reviews of instructional designs from a constructivist perspective (e.g., Dean, Hubbell, Pitler & Stone, 2012; Frey, 2010; Noonan, 2013). Lastly, there are reviews of instructional practices from a cognitive science perspective (e.g., Mayer, 2010; Howard-Jones, 2010; Jensen, 2005; Wolfe, 2010). As a review of individual papers that explore a singular instructional practice exceeds the scope of this chapter, what follows are a sampling of reviews of best instructional practices from the constructivist perspective and then from a cognitive science-based perspective.

Constructivist Instructional Design

In spring 1991, the Association for Supervision and Curriculum Development assembled an advisory panel on improving student achievement. The panel concluded that teachers needed a wide range of effective instructional tools to promote teaching that was relevant, multicultural, engaging, and appealing to learners with diverse learning styles. Derived from an extensive review of research, these instructional tools served as the basis for the iteration of 16 strategies that were found to be effective in promoting learning and engagement. These strategies were organized under a more broad framework as follows: strategies that capitalize on students' strengths; strategies that match instructional methods to student instructional needs; strategies that increase motivation, interest, and engagement; strategies that create a variety of learning configurations; and strategies that make connections for understanding (Cole, 2008).

Another example of constructivist-based instructional techniques can be drawn from good practice literature, including books, training manuals, web sites, workshops, and journals, as reviewed by Barkley (2010). Barkley's Student Engagement Techniques (SETs) are 50 "field tested" learning activities found effective in engaging undergraduate students that can be organized into two main categories: techniques to engage students in the content of a course; and techniques for developing attitudes, values, and self-awareness of students (Barkley, 2010). Examples of these techniques include employing constructs of split-room debates, small group tutorials, role-play, Think-Aloud-Pair-Problem Solving, case studies, dyadic interviews, learning logs, student generated rubrics, and triad listening (Barkley, 2010).

Cognitive Science Instructional Design

In terms of the cognitive science-based perspective, *Teaching with the Brain in Mind* (Jensen, 2005) and the *Handbook of Research on Learning and Instruction* (Mayer & Alexander, 2010) provide a thorough examination and review of instructional designs based in the research of cognitive psychology and brain research.

Jensen (2005) identifies five necessary steps to effective instruction: engagement, framing, acquisition, elaboration, and memory strengthening. Engagement includes creating a positive social climate and using journaling, humor, art, group rituals, activities, affirmations, and stretching as the setup to actual instruction. Framing, which activates neuronal assemblies, is a tool that creates an intentional bias toward what follows. This can include a picture, background activity, or any other construct that would 'hook' the

learner mentally. Acquisition is the step that includes cooperative or collaborative learning, an activity, or experiences that focus on the input of the learner. Elaboration is determining whether learners have developed a deep understanding of material, something that can be evaluated through peer editing, feedback, competitions, and partner quizzes. Lastly, memory strengthening draws on the principle that learners will recall learned material more in the first hour following a learning experience than in the days that follow. As such, Jensen suggests having learners share their understanding of content with partners, using drama, creating acronyms, visual representations, rhymes, quizzes, or mental models as a way to reinforce and encapsulate learning in a format that facilitates ease of recall.

Finally, the *Handbook of Research on Learning and Instruction* (2010) is a collection of papers that focus on the following strategies that influence student achievement: feedback; well-crafted, well-positioned examples; self-explanations; peer interaction; inquiry-based instruction; discussion; computer-based media; tutoring; and visual-spatial representations and visualizations. Although specific strategies are mentioned in this text, the emphasis in this handbook is oriented toward a broad explanation of instructional design clusters – the specific iterances in the classroom still dependent on a teacher's assessment of what strategies ultimately to deploy to promote engagement and learning.

Summation for Enhancing Learner Engagement

In the review of best instructional designs in the classroom, the constructivist and cognitive science perspectives essentially both advocate for two guiding principles to promote engagement and learning. The first principle recognizes that superior instruction promotes an interactive learning experience in a dialogic paradigm that features a combination of learner and teacher, learner with learner(s), or learner with content. The second principle is that the choice of strategies to support content retention, transfer, and recall is largely dependent on the nature of the content, the educational aims of the teacher, and the needs of the individual learner. These two principles echo Dewey's (1929) engineering metaphor of the art and science of education and instructional design:

> There is a science of bridge building in the sense that there is a certain body of *independent* scientific material, say mathematics and mechanics, from which selections may be made and the selections organized to bring about more effective solution in practice of the difficulties and obstructions that present themselves in actual building of bridges. It is the way the material is handled and organized with reference to a purpose that gives us a right to speak of a science of bridge   building, although the building itself is an art, not a science. The sciences of mechanics and mathematics are, in themselves, the science, which they are, not sciences of bridge building. They *become* the latter when selected portions of them are focused upon the problems presented in the art of bridge building. (pp. 34-35).

Thus, best instructional practices are made manifest when the science of the material, or the science of how a learner learns, informs the selection, or art, of instructional designs that best construct the learning experiences in the classroom.

Essentially, this informed process of selection facilitates the engagement of the learner by directing their focus to the salient issues of the content under consideration. In this way, the reciprocity of science and art give rise to superior instructional practices in the classroom – the employment of which defies formulaic application but requires a dynamic approach that embraces the phenomenology of engagement: the strengths and needs of individual learner, the nature and scope of the content, and the delivery of that content that promotes the attentional focus of, and meaning making by, the learner.

Adaptive Instructional Strategies to Enhance Learner Grit

Research has increasingly focused on the study of non-cognitive factors that contribute to overall success in academic environments. Studies showing the positive correlation between intelligence and achievement are vast and well documented; however, the effect that non-cognitive skills have on achievement is gaining interest as a valid predictor of accomplishment. The following section of this chapter explores the specific non-cognitive factor of grit and the related terms perseverance and tenacity and their relationship to learning success and overall academic progress. We then relate these factors to instructional strategies that may be useful in intelligent tutoring systems, specifically as they apply to design and capabilities of GIFT.

Grit

Grit is a relatively new term but not a new concept in the study of factors that contribute to achievement and success. Duckworth, Peterson, Matthews, and Kelly (2007) first introduced the term grit in their studies of predicting achievement and success among high-achieving individuals and specifically identify grit as the perseverance and passion for long-term goals. They describe gritty individuals as those who deliberately set high goals and pursue them over long periods of time despite occasional setbacks and lack of positive feedback (Duckworth et al., 2007). Their research pursued the study of grit through six distinct experimental environments using a self-report questionnaire they developed called The Grit Scale.

Initially in 2004, registered users of the University of Pennsylvania's Department of Psychology's web site www.authentichappiness.org were invited to validate a 27-point survey developed to measure grit that was posted on the web site. Questions were designed to be valid for both adult and adolescent users, not relate specifically to work or school, and address one's overall experience with maintaining long-term projects despite adversity or lack of immediate reward. By October 2005, they had collected data on 1545 adults age older than 25 years – mean age 45 years (73% female; 27% male). After analysis, 15 items were removed resulting in a 12-item questionnaire, which was validated as a consistent measure of grit – 6 items pertaining to the consistency of interests and 6 items related to persistence of effort.

The first cross-sectional study was designed to validate the grit scale. Participants were asked to answer the grit questionnaire and provide their age and highest level of education. Results revealed persons with more education also had more grit than persons with less education of the same age. In addition, when controlled for age, post college graduates had the most grit and persons with Associate's degrees had more grit than persons with less education. When controlled for education, it was suggested that grit tends to increase with age but the study was unable to reach a reliable conclusion. Although all answers in Study 1 were self-reported and may be subject to social desirability bias, it seems conclusive that attainment of higher educational levels is associated with increased levels of grit.

The five additional studies using the 12-point scale developed by Duckworth et al. (2007) were also successful in associating increased levels of accomplishment with increased levels of grit. Study 2 investigated the effect of conscientiousness and other Big Five traits to determine if grit had better predictive validity than the Big Five traits. In addition to the 12-point grit scale, the study also included the Big Five Inventory Questionnaire and specifically asked how often those surveyed changed careers to determine if grittier individuals had fewer career changes. Results from 690 participants – mean age 45 years (80% women; 20% men) – who answered the study on the same web site as in Study 1 revealed that grit was a better predictor than conscientiousness when studied for education and age. Persons with less education had less grit – highest scorers of grit were persons with advanced degrees and persons with Associates degrees followed by persons with Bachelor's degrees. Persons with some college had the least amount of grit. In addition, Study 2 also correlated long-term career stability with increased grit.

Study 3 explored the relationship between grit and the GPA scores of undergraduate students majoring in psychology at the highly selective University of Pennsylvania. Participants were recruited via email sent to 390 students and resulted in 139 participants – 69% female; 31% male whose combined average SAT score was 1415. In addition to the 12-item grit scale, they were also asked to report their gender, SAT scores, current GPAs, and expected year of graduation. Results showed grittier persons had higher GPAs especially when SAT scores were held constant. It also revealed that those with lower SAT scores had higher grit than those who scored higher on the SAT, suggesting that grittier persons make up for lack of intelligence by working harder (Duckworth et al., 2007).

Study 4 examined the effect of grit on summer retention and the following year's GPAs in 1218 freshman cadets at the rigorous United States Military Academy at West Point in 2004. Incoming students completed the 12-item grit scale and the Brief Self Control Scale (BSCS) (Tangney, Baumeister & Boone, 2004) within 2 to 3 days of arrival at West Point. The study also compared the Whole Candidate Score, an internal scoring device used by West Point that contains analysis of SAT scores, high school class rank, participation in extracurricular activities, and a physical endurance score. Results were compared following the grueling summer training program called Beast Barracks and showed grit did not affect the Whole Candidate Score but did compare with self-control as determined by answers on the BSCS. Retention following the summer training was most predicted by grit. The Whole Candidate Score, however, best predicted GPAs. GPAs were also more strongly predicted by measure of self-control than by grit. Duckworth et al. (2007) also measured self control separately from grit due to different levels of perseverance necessary to achieve short-term versus long-term goals.

Study 5 was a replication of Study 4 to determine the predictive validity of grit over Big Five Conscientiousness. The 12-item grit scale and 9-item Conscientiousness subscale of the Big Five Inventory were administered to 1308 incoming freshman one day after arriving at the United States Military Academy at West Point in 2006. Results showed summer retention was better predicted by grit than by either conscientiousness or Whole Candidate Score.

Study 6 was a longitudinal study conducted on 175 finalists of the Scripps National Spelling Bee in 2006 to test the hypotheses that grit better predicted time of study and multiple final round appearances. The group consisted of 48% female; 52% male with a mean age of 13.20 years. The 12-item grit scale, the BCBS and the Verbal IQ test, a Similarities subtest of the Weschler Intelligence Scale for Children – III (Weschler, 1991) were administered to all participants prior to or after the competition. Students were also asked to report how many hours per day they studied for the finals during the week and how many hours per day they studied over the weekend. Grit predicted advancement to higher rounds and showed that grittier individuals studied longer. The study also revealed grit to have a strong relationship to self-control. Duckworth et al. (2007) have shown that IQ is not the sole predictor of success through their series of six studies.

Academic Tenacity

Many others have also questioned what traits make some more successful than peers of equal intelligence. Weschler was a great advocate of including non-intellective factors in intelligence testing and, in 1943, agreed with the earlier work by Alexander (1935) that drive, persistence, and interest were considerably underaccounted for in measures of intelligence testing (Weschler, 1943). More recently, Dweck, Walton, and Cohen (2011) identify the term "academic tenacity" as a mindset to focus on longer-term goals and persevere through short-term challenges to achieve those goals. An academic tenacious student, according to this research, typically sees success in education as a means to a longer-term goal, is able to self-regulate time and attention, and readily accepts challenge as an opportunity to learn new things (Dweck et al., 2011).

According to Dweck et al. (2011), improving a student's mindset toward learning will help motivate the student and encourage habits that lead to success in educational pursuits. Studies of lower income, racially diverse student populations over two years revealed two distinct mindsets contributed to success or failure in academic environments (Blackwell, Trzesniewski & Dweck, 2007, Study 1). Students with a "fixed" mindset believe intelligence is static and little can be done to improve it. These types of students are overly concerned with their ability academically and shy away from tasks at which they do not excel. They interpret failure as humiliating due to what they perceive as their intellectual inadequacies and are more prone to give up. They are more concerned with their performance than learning.

In contrast, students with a "growth" mindset believe intelligence is expandable and can increase with study and dedication toward learning. They believe failure is an opportunity to learn and is not due to intellectual limits. They seek the opportunity to learn in all endeavors and relish challenge (Dweck & Leggett, 1988). Blackwell et al. (2007) showed that although all students entered the 7th grade with similar performance scores, math grades in those with a fixed mindset decreased while math grades in those with a growth mindset increased over a two-year span. The students with a growth mindset "earned higher grades because they valued learning over looking smart" (Blackwell et al., 2007). Academic tenacity they concluded requires a growth mindset.

Academic Perseverance

Farrington, Roderick, Allensworth, Nagaoka, Keyes, Johnson, and Beechum, (2012) have studied non-cognitive factors that include strategies, attitudes, and behaviors in addition to non-cognitive skills that that affect academic achievement. They posit that student academic behaviors have the greatest effect on learning success and grades (Farrington et al., 2012). In their review, Farrington et al. (2012) developed five categories of non-cognitive factors that are related to academic success: academic behaviors; academic perseverance; academic mindsets; learning strategies; and social skills. The term "academic perseverance" describes students who "behave in an engaged, focused, and persistent manner in pursuit of academic goals, despite obstacles, setbacks, and distractions" (Farrington et al., 2012, p. 20). Students who have academic perseverance achieve academically by continuing to try to get a good grade in a challenging class despite failing performances on tests and would continue to try to understand difficult material without giving up (Farrington et al., 2012, p. 20). The review focuses on grit and self-control, which is the ability to forego immediate temptations for the sake of a less tangible goal.

The role of grit on academic performance as seen through Duckworth's grit study of SAT scores of 139 University of Pennsylvania students (Duckworth et al., 2007) is discussed. Although Farrington et al. (2012) agree that students with more grit may indeed achieve more, they argue that in Duckworth et al. (2007) specific studies, the participants were so uniquely homogeneous due to their high SAT scores, the findings that those with lower SAT scores are more gritty to overcome academic short fallings may not necessarily be valid in studies involving more heterogeneous populations. Additionally, Duckworth et al. (2007) studies consider grit to be an inherent personality trait. Farrington et al. (2012) consider academic perseverance a malleable behavior that can be changed to increase academic achievement. Although overall, Farrington et al. (2012) agree that the studies of Duckworth et al. (2007) show a relationship between grit and academic perseverance, they encourage more research to determine a more causal-related relationship.

Dweck et al.'s (2011) academic tenacity is also discussed. Although Farrington et al. (2012) believe that the factors of mindset, academic skills, learning strategies, and personality included in academic tenacity may contribute in total to academic perseverance, for the purpose of their analysis, they chose to leave those factors out and focus solely on the measure of academic performance without the variables that may influence it (p. 20).

Academic perseverance also includes what the researchers term "effortful control," meaning the inclusion of self-control and delayed gratification. Farrington et al. (2012) conclude the role of perseverance on academic success is vague and it is more effective to focus on teaching positive academic mindsets and learning strategies that have been shown to improve academic performance over time (pp. 26, 27).

Grit, Tenacity, and Perseverance

In 2013, the U.S. Office of Educational Technology released a draft report entitled, "Promoting Grit, Tenacity, and Perseverance: Critical Factors for Success in the 21st Century" (Shechtman, DeBarger, Dornsife, Rosier & Yarnall, 2013). For the purpose of the report, they define grit as the "perseverance to accomplish long-term or higher-order goals in the face of challenges and setbacks, engaging the student's psychological resources, such as their academic mindsets, effortful control, and strategies and tactics" (Shechtman et al., 2013, p. 15). The authors believe non-cognitive factors are malleable and can be taught effectively to students to increase academic performance. The report agrees with Dweck et al. (2007) that growth mindsets are essential to facilitate perseverance and academic achievement. Csikszentmihalyi's (1990) "Flow" is also supported to develop a sense of perseverance in which students are best challenged and motivated to learn by material that is within or only moderately beyond their skill level – material that is too easy induces boredom; material that is too difficult creates frustration.

Presenting learning material in a way that is relevant or correlates with students' personal goals and interests has also been shown to be effective in promoting perseverance. Basic methods such as choosing material and independently defining timelines and approaches to getting work accomplished has also resulted in increased engagement from students (Shechtman et al., 2013).

Relevant to the review, associated methods proposed to increase student grit, tenacity, and perseverance include instructional emphasis on mindsets, learning strategies, and resilience through the use of "research-based best practices" that include the use of technology. They stress that technological approaches should be developed by interdisciplinary teams that are experts in the learning sciences, software design, and domain-specific content (Shechtman et al., 2013, p. xiii). In addition, new technology-based digital learning environments that support educational data mining and affective computing should be integrated to detect potential academic vulnerabilities in students and potentially steer them back before academic failure. More research is required to determine the costs and benefits of promoting grit in different learning environments and identifying potential situations where grit may be detrimental.

Learning Strategies to Enhance Learner Grit, Tenacity, and Perseverance

Review of the literature shows gritty, academically tenacious, perseverant learners have strong desire and motivation to learn. They tend to have a growth mindset about learning and are not likely to give up or be dissuaded due to failure. They are also independent and would be more likely to embrace the opportunity to have some choice over their instruction. Several instructional strategies that support these character traits are in the literature. No studies that specifically address learner grit in relation to instruction and learning have been found. However, the following adaptive instructional research supports traits consistent with gritty, tenacious, and perseverant learners.

Goal-Directed Feedback

Goal-directed feedback may be effective for gritty learners since it does not specifically supply the answer, but addresses whether the step contributes to achievement of the goal or not (Shute, 2007, p. 12): "One way to influence learners' goal orientations (e.g., to shift from a focus on performing to an emphasis on learning) is via formative feedback. Hoska (1993) showed how goal-orientation feedback can modify a

learner's view of intelligence, by helping a learner see that (a) ability and skill can be developed through practice, (b) effort is critical to increasing this skill, and (c) mistakes are part of the skill-acquisition process" (Shute, 2007, p.13).

Self-Explanation and Self-Examination

Self-explanation (Chi et al., 1994) and self-examination are also valid strategies that produce greater learning with computer-based tutoring systems. It has been shown that learners benefit from reviewing their answer choices based on why they selected that answer and if they have enough information to reach a specific conclusion to correctly select an answer (Chi et al., 1994; Durlach et al., 2011). Self-explanation has been used in AutoTutor as a natural language dialog with the computer and the student (VanLehn, Graesser, Jackson, Jordan, Olney & Rosé, 2007). Self-examination forces the student to review their work and determine at what point an error was made in a problem and attempt to correct the wrong step. The prompt may not occur immediately when the error was committed and help is provided only after the student has difficulty finding or fixing an error. This strategy of feedback is also called self-correction.

Summation for Enhancing Learner Grit

Tailoring of instruction depends on what is being taught or learned and the individual learner characteristics, such as motivation, interest, and ability level for that subject matter. Students should be assessed to determine their current levels of ability, motivation, and interest, and should be re-administered throughout the learning activity to continuously assess their levels and adapt instruction. GIFT effectively assesses the learner through its ability to author surveys and interpret physiologic data or assess historical data.

For students who are perceived as more academically proficient, motivated, and determined, several instructional strategies can be more effective than others. Through the pedagogical module in GIFT, instruction can be adapted based on a learner's historic and real-time data.

Much of the research on non-cognitive factors focuses on correlation rather than causal relationships (e.g., presence of grit and high educational achievement and not if increasing grit will increase educational achievement [Farrington, 2012, p. 13]). It is difficult to translate study results into direct classroom or computer-based methods that produce reliable results.

The conclusion on adaptive instructional strategies for gritty learners is largely that it depends. It is situational. Although many adaptive instructional strategies have been used and studied, no single adaptive instructional strategy has proven to be more reliable across all domain environments than another. Some studies used only one adaptive technique; some used as many as three. However, no one stands out as the across-the-board superior adaptive strategy in all cases for all students.

Adaptive instruction via a computer-based tutor is in its infancy in many ways – many studies have been conducted and results obtained; however, similar to human learning via any method, it is difficult to determine precisely which adaptive strategy consistently outperforms another in all situations. Review of the current literature regarding adaptive instructional strategies for enhancing learner grit shows that it continues to depend on individual learner traits (e.g., values, motivation, goals) and can vary even within a single individual over time and is situational. Error-based feedback; repetitive, spaced strategy; mastery learning; faded-worked examples; and metacognitive prompting all are effective for gritty persons, and all persons for that matter, and should continue to be studied and for now, included in adaptive, computer-based tutoring.

Although significant advances have occurred in adaptive instruction via a computer-based tutor, the results of any given instruction still depend on the specific learner, how and when the adaptation is presented; and the learning environment itself. Knowing if a learner has traits associated with grit, tenacity, and perseverance will help determine useful instructional strategies. For the present, the experts will continue to study the learners and the learners will continue to be studied. Any improvement in learning of any sort should be considered successful.

Conclusions and Recommendations

We provided a review of instructional techniques, strategies, and tactics that influence learning, affect, engagement, and grit. We distinguished instructional techniques as domain-independent and, while largely learner-independent in their application, they may be tailored to support adaptive tutoring. We noted instructional strategies as domain-independent but learner-dependent, and instructional tactics as domain- and learner-dependent. Techniques were generally easier to implement.

A set of instructional techniques were reviewed prior to evaluating instructional strategies and tactics in the literature, which influence affect, engagement, and grit. In addition to their effect on learning, affect, engagement, and grit, we also examined their influence in developing other desirable traits and their ease of implementation (e.g., authoring and reusability).

In the area of affect, we reviewed three areas: developing emotional intelligence in adaptive tutoring systems; modeling the response of expert human tutors to mange affect; and using the ZPD to manage affect. Developing emotional intelligence in ITSs is considered moderate to difficult depending on the goals set for the system. The influence of some behaviors represented in the INSPIRE model may not be mutually exclusive (e.g., nurturant and indirect), and the implementation may be subject to wide variability due to how frequently each behavior/feedback is triggered during tutoring. Any strategy that might be overused is undesirable.

As part of our review of the response of expert human tutors, we also evaluated Anderson et al.'s (1987) principles later elaborated by Corbett et al. (1997) with respect to their ease of implementation and many of them were found to be relatively easy to implement in an architecture like GIFT. While ZPD is mentioned often in the tutoring literature, relatively few models have been implemented and only one, Murray and Arroyo (2002) was operationalized for use across tutoring domains.

Next we evaluated engagement methods and specifically focused on the nexus of engagement, orientation, and attention in instructional design. Several strategies/tactics/techniques were noted and we chose to focus on instructional design from a constructivist and cognitive science perspective. Omitted from this analysis were the instructional designs influenced from a behavioralist epistemological framework and instructional designs based on the information-processing framework, as this latter school of thought has largely been incorporated under the current epistemological framework of cognitive science more generally. Instructional design principles of note included promoting interactive learning; selecting strategies based upon scaffolding instruction to consider cognitive load; supporting long-term memory transfer through personal meaning making; selecting strategies based upon how the learner prefers to learn; constructing knowledge by linking the learner's prior knowledge to new content; and selecting content and delivery methods that promote attentional focus.

Lastly, we examined instructional strategies to enhance learner grit or perseverance. The results of six studies were discussed along with a comparison of grit with academic tenacity and academic perseverance. We concluded by examining three methods that may be effective for gritty learners: goal-directed feedback, self-explanation, and self-examination.

The strengths of the instructional strategy methods reviewed here are based on the notion that they stem either from empirically validated educational research or field reports of successful temporal classroom instruction, including the one-to-one tutoring paradigm. Although the information regarding successful instructional paradigms is vast and is certainly not exhaustively reported here, this review provides the reader with sources from which to model and apply what we consider as best instructional temporal practices that can inform instructional design in a simulated platform.

Specific to GIFT, these sources of instructional strategies can be used to help shape GIFT's Domain Module, most importantly in examining how content is shaped and delivered to optimize the learning experiences of trainees. This optimization includes targeting mastery learning of domain-specific skills as well as broader analogical and problem-solving thinking skills across a variety of content-specific domains. Further, it is the hope that this review will help inform instructional designs in the Pedagogical Module, particularly in evaluating how to most effectively construct feedback features that will speak to promoting trainee grit and perseverance, with the additional consideration of the emotional, cognitive, and behavioral elements that may shape trainee engagement.

References

Ainley, M. (2012). Students' interest and engagement in classroom activities. In S. Christenson, A., Reschly & C.Wylie, (Eds.), *Handbook of research on student engagement*, (283-302). New York, NY: Springer.

Alexander, S., Sarrafzadeh, A. & Fan, C. (2003). Pay Attention! The Computer is Watching: Affective Tutoring Systems. In G. Richards (Ed.), Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2003 (pp. 1463-1466). Chesapeake, VA: AACE.

Alexander, W. P. (1935). Intelligence, concrete and abstract. *British Journal of Psychology*. Monograph Supplement No. 19.

American Psychological Association (2007). VandenBos, Gary R., ed. *APA Dictionary of Psychology* Washington, DC: American Psychological Association, page 26.

Anderson, J., Boyle, C., Farrell, R. & Reiser, B. (1987). Cognitive principles in the design of computer tutors. In P. Morris (Ed.), *Modeling cognition*. NY: John Wiley.

Anderson, J. R., Corbett, A. T., Koedinger, K. R. & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167-207.

Andre, E., Rehm, M., Minker, W. & Bühler, D. (2004). Endowing spoken language dialogue systems with emotional intelligence. In *Affective Dialogue Systems* (pp. 178-187). Springer Berlin Heidelberg.

Atkinson, R. K., Renkl, A. and Merrill, M.M. (2003). "Transitioning From Studying Examples to Solving Problems: Effects of Self-Explanation Prompts and Fading Worked-Out Steps." Journal of Educational Psychology 95(4)

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52* , 1–26.

Barkley, E. (2010). *Student engagement techniques: A handbook for college faculty.* Hoboken, NJ: Jossey-Bass.

Blackwell, L., Trzesniewski, K.H. & Dweck, C.S. (2007). Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, *78*(1), 246 – 263.

Blumenfeld, P. C., Kempler, T. M. & Krajcik, J. S. (2006). Motivation and Cognitive Engagement in Learning Environments.

Bower, G.H. (1981). Mood and memory. American Psychologist. 1981 Feb Vol 36(2) 129-148.

Bower, G. H. & Forgas, J. P. (2000). Affect, memory, and social cognition. In E. Eich, J. F. Kihlstrom, G. H. Bower, J. P. Forgas & P. M. Niedenthal (Eds.), Cognition and emotion (pp. 87–168). New York: Oxford University Press.

Bucciarelli, M. (2004). How the construction of mental models improve learning. *Mind & Society* 6, 67-89. DOI 10.1007/s11299-006-0026-y

Byrnes, J.P. (1996). Cognitive Development and Learning in Instructional Contexts. Boston: Allyn and Bacon.

Calvo, R. A. & D'Mello, S. K. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1, 18-37.

Chan, M.S. and Black, J.B. (2006) Direct-manipulation animation: Incorporating the haptic channel in the learning process to support middle school students in science learning and mental model acquisition. *Proceedings of International Conference of the Learning Sciences*. Mahwah, NJ: LEA.

Chandler, P. & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293-332.

Chi, M. T., H., de Leeuw, N., Chui, M. H. & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439-477.

Clark, R.C., Nguyen, F., and Sweller, J. (2006). Efficiency in learning: evidence-based guidelines to manage cognitive load. Wiley & Sons. ISBN 0-7879-7728-4.

Cole, R. (Ed.) (2008). *Educating everybody's children: Diverse teaching strategies for diverse learners*. Alexandria, VA: Association for Supervision & Curriculum Development (ASCD).

Corbett A. T., Koedinger, K. R. & Anderson, J. R. (1997). Intelligent tutoring systems. In M. G. Helander, T. K. Landauer & P. V. Prabhu (Eds.), *Handbook of human-computer interaction* (pp. 849–874). Amsterdam: Elsevier.

Corno, L. & Mandinach, E. B. (1983). The role of cognitive engagement in classroom learning and motivation. *Educational psychologist*, *18*(2), 88-108.

Corrigan, L., Peters, C. and Castellano, G. (2013). Identifying Task Engagement: Toward Personalised Interactions with Educational Robots. 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII) IEEE, 2013, ISSN: 2156-8103.

Csikszentmihalyi, M. (1990). Flow: the psychology of optimal experience. Harper Perennial, New York, NY.

D'Mello, S. & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145-157.

Dean, C., Hubbell, E., Pitler, H. & Stone, B. (2012). *Classroom Instruction that Works* Alexandria, VA: Association for Supervision and Curriculum.

Dempster, F.N. (1988). The spacing effect. *American Psychologist*, 43, 627-634.

de Quervain, D., Roozendaal, B., Nitsch, R., McGaugh, J. & Hock, C. (200). Acute cortisone administration impairs retrieval of long-term declarative memory in humans, *Nature Neuroscience*, 3, 313-314.

Dewey, J. (1929). *The sources of a science of education*. New York, NY: Liveright.

Dewey, J. (1944). *Democracy and Education*. New York, NY: The Free Press.

Duckworth, A.L., Peterson, C., Matthews, M.D. & Kelly, D.R. (2007). Grit: Perseverance and Passion for Long-term Goals. *Journal of Personality and Social Psychology*, *92*(6), 1087-1101.

Durlach, P. & Ray, J. (2011). Designing Adaptive Instructional Environments: Insights from Empirical Evidence. *Technical Report 1297.* Arlington, VA: United States Army Research Institute For the Behavioral and Social Sciences.

Dweck, C. S. & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.

Dweck, C.S., Walton, G.M. & Cohen, G.L. (2011). Academic Tenacity: Mindsets and Skills that Promote Long-term Learning. Paper prepared for the Gates Foundation. Seattle, WA: Bill & Melinda Gates Foundation.

Eldridge, M. (1998). *Transforming experience: John Dewey's cultural instrumentalism*. Nashville: Vanderbilt University Press.

Farrington, C.A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T.S., Johnson, D.W. & Beechum, N.O. (2012). *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance: A Critical Literature Review*. Chicago, IL: The University of Chicago Consortium on Chicago School Research.

Fletcher, J.D. and Sottilare, R. (2013). Shared Mental Models and Intelligent Tutoring for Teams. In R. Sottilare, A. Graesser, X. Hu, and H. Holden (Eds.) *Design Recommendations for Intelligent Tutoring Systems: Volume I - Learner Modeling*. Army Research Laboratory.

Fogarty, R. (1999). Architects of the intellect. *Educational Leadership.* 57 (3), 76-78.

Fredricks, J., Blumenfeld, P. & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74,* 59-109.

Frey, N. (2010). *Teaching Practices that Work, Effective Teacher's Guide: 50 Ways toEngage Students and Promote Interactive Learning*. New York, NY: Guildford Press.

Friedman, R., Fishback, A., Forster, J. & Werth, L. (2003). Attentional priming effects on creativity. *Creativity Research Journal*, 15 (2), 277-286.

Gebhard, P., (2005). ALMA - A Layered Model of Affect. In Proceedings of the *Fourth International Joint Conference on Autonomous Agents & Multi Agent Systems* (AAMAS), Utrecht, Netherlands, July 25 to 29, 2005.

Gold, P. and van Buskirk R. (1975). Facilitation of time-dependent memory processes with posttrial epinephrine injections. Behavioral Biology, 13, 145-153.

Goleman, D., (1995). Emotional Intelligence, New York, NY, England: Bantam Books, Inc.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A. & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments & Computers, 36*(2), 180-192.

Hernandez, Y., Noguez, J., Sucar, E. and Arroyo-Figueroa G. (2006). Incorporating an Affective Model to an Intelligent Tutor for Mobile Robotics. In Proceedings of the 36th Annual Frontiers in Education Conference, San Diego, California, USA.

Hickman, L., Neubert, S. & Reich, K. (Eds.) (2009). *John Dewey between pragmatism and constructivism*. New York, NY: Fordham University Press.

Howard-Jones, P., Blakemore, S., Samuel, E., Summers, I. & Claxton, G. (2005). Semantic divergence and creative story generation: An fMRI investigation. Cognitive Brain Research, 25, 240-250.

Howard-Jones, P. (2010). *Introducing neuroeducational research: Neuroscience, education and the brain from contexts to practice.* New York, NY: Routledge.

Hoffmann, L. and Krämer, N.C. (2011). "How Should an Artificial Entity be Embodied? Comparing the Effects of a Physically Present Robot and its Virtual Representation," in Proceedings of *Workshop on Social Robotic Telepresence, Human-Robotic Interaction 2011*, Lausanne, Switzerland, 2011

Hoska, D. M. (1993). Motivating learners through CBI feedback: Developing a positive learner perspective. In V. Dempsey & G. C. Sales (Eds.), *Interactive Instruction and Feedback* (pp. 105–132). Englewood Cliffs, N.J.: Educational Technology Publications.

Immordino-Yang, M. & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain and Education*, 1, 3-10.

Isen, A.M. (2000). Positive Affect and Decision Making, in M.Lewis & J.M. Haviland-Jones (eds.), Handbook of Emotions, second edition, The Guildford Press, 417- 435, 2000.

Isen, A. M. (2003). Positive Affect as a Source of Human Strength. In L. Aspinwall & U. Staudinger (Eds.), A Psychology of Human Strengths (pp. 179-195). The American Psychological Association, 2003, Washing-ton, D.C.

Isen, A. (2004). Some Perspectives on Positive Feelings and Emotions: Positive Affect Facilitates Thinking and Problem Solving. In N. F. a. A. F. Antony Manstead (Ed.), The Amsterdam Symposium. New York: Cam-bridge.

Isen, A. M., Daubman, K. A. & Nowicki, G. P. (1987). Positive affect facilitates creative problem solving. Journal of Personality and Social Psychology, 52(6), 1122-1131.

Isen, A. M., and Erez, A. (2006). Some measurement issues in the study of affect. In A. Ong & M. vanDulman (eds.), Oxford Handbook of Methods in Positive Psychology. Oxford University Press, New York, USA.

Jensen, E. (2005). *Teaching with the Brain in Mind*. Alexandria, VA: Association for Supervision & Curriculum Development (ASCD).

Johnson, A. and Taatgen, N. (2005). In R. Proctor & K. Vu (Eds.) Handbook of Human Factors in Web Design, Chapter 25 - User Modeling. New Jersey: Lawrence Erlbaum Associates, Inc.

Jordan, A., Orison, C. & Stack, A. (2008). *Approaches to Learning: A Guide for Teachers*. Berkshire, England: Open University Press.

Kearsley, G. & Shneiderman, B. (1998). Engagement Theory: A Framework for Technology-Based Teaching and Learning. *Educational technology*, *38*(5), 20-23.

Klein, G. & Baxter, H. (2006). Cognitive Transformation Theory: Contrasting Cognitive and Behavioral Learning. Presented at the Interservice/Industry Training Systems and Education Conference, Orlando, Florida, De-cember 2006.

Klesen, M. (2002). Report on affective reasoning and cultural diversity. NECA: A Net Environment for Embodied Emotional Conversational Agents.

Kounis, J., Fleck, J., Geen, D., Payne, L., Stevenson, J., Bowden, E. & Jung-Beeman, M. (2008). The origins of insight in resting-state brain activity. *Neuropsychologia*, 46, 281-291.

Kuhn, D. & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science* 22(4), 545-552.

Lane, H.C. & Johnson, W.L. (2008). Intelligent Tutoring and Pedagogical Experience Manipulation in Virtual Learning Environments. In J. Cohn, D. Nicholson & D. Schmorrow (Eds.), *The PSI Handbook of Virtual Environments for Training and Education* (pp. 393-406). Westport, CT: Praeger Security International.

Lepper, M. R., Drake, M. & O'Donnell-Johnson, T. M. (1997). Scaffolding techniques of expert human tutors. In K. Hogan & M. Pressley (Eds), *Scaffolding student learning: Instructional approaches and issues* (pp. 108-144). New York: Brookline Books.

Lepper, M. and Woolverton, M. (2002). The Wisdom of Practice: Lessons Learned from the Study of Highly Effective Tutors. In J. Aronson (Ed.), Improving academic achievement: impact of psychological factors on education (pp. 135-158). New York: Academic Press.

Linnenbrink, E. A., and Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review*, 31, 313-327.

Mayer, R. (2010). Instruction based on visualization. In R. Mayer & P. Alexander, (Eds.), *Handbook of research on learning and instruction,* (427-445)*.* Florence, KY: Routledge.

Mayer, R. & Alexander, P. (Eds.) (2010). *Handbook of research on learning and instruction*. Florence, KY: Routledge.

Merrill, D. , Reiser, B, Ranney, M., and Trafton, J. (1992). Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. The Journal of the Learning Sciences, 2(3), 277-305.

Murray, T. & Arroyo, I. (2002). Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems. Sixth International Conference on Intelligent Tutoring Systems. Springer.

Noonan, S. (2013). *How real teachers learn to engage all learners*. Lanham, MD: R&L   Education.

Norman, D. (2002). Emotion and Design: Attractive Things Work Better. Interactions, 9, 4, 36-42.

Nye, B., Sottilare, R., Ragusa, C. & Hoffman, M. (2014, in press). Prologue - Defining Instructional Challenges, Strategies, and Tactics for Adaptive Intelligent Tutoring Systems. In R. Sottilare, A. Graesser, X. Hu & B. Goldberg (Eds.) *Design Recommendations for Intelligent Tutoring Systems: Volume 2 - Instructional Management.* Army Research Laboratory, Orlando, Florida. ISBN: 978-0-9893923-3.

Pea, R. D. (1988). Putting knowledge to use. In Raymond S. Nickerson & Philip R. Zodhiates (Eds.), *Technology in education: Looking toward 2020*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Person, N. K., Kreuz, R. J., Zwaan, R. A. & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction*, 13(2), 161−188.

Picard, R. (2006). Building an Affective Learning Companion. Keynote address at the 8th International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan. Retrieved from http://www.its2006.org/ITS_keynote/ITS2006_01.pdf

Pritchard, A. & Woolard, J. (2010). *Psychology for the Classroom: Constructivism and Social Learning*. Florence, KY: Routledge.

Pugh, K. J., Linnenbrink-Garcia, L., Koskey, K. L., Stewart, V. C. & Manzey, C. (2010). Motivation, learning, and transformative experience: A study of deep engagement in science. *Science Education*, *94*(1), 1-28.

Reyes, M., Brackett, M., Rivers, S., White, M. & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of   Educational Psychology* 104(3), 700-712.

Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192. Retrieved on 20 December 2013 from http://ezproxy.cul.columbia.edu/login?url=http://search.proquest.com/docview/198832146?accountid=10226

Rodrigues, M., Novais, P., and Santos, M.F. (2005). Future Challenges in Intelligent Tutoring Systems − A Framework. In Proceedings of the Third International Conference on Multimedia and Information & Communication Technologies in Education, Cáceres, Spain.

Rosenberg-Kima, R.B., Plant, E.A., Baylor, A.L. and Doerr, C.E. (2007). Changing Attitudes and Performance with Computer-Generated Social Models. In Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED), p. 51-58. Marina del Rey, CA.

Schunk, D. & Mullen, C. (2012). Self-efficacy as an engaged learner. In S. Christenson,   A., Reschly & C.Wylie, (Eds.), *Handbook of research on student engagement*, (219-235). New York, NY: Springer.

Schwartz, D., Chase, C., Oppezzo, M. & Chin, D. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology* 103(4) 759-775.

Sessink, O., Beeftink, H., Tramper, J. & Hartog, R. (2007). Proteus: A Lecturer-Friendly Adaptive Tutoring System. Journal of Interactive Learning Research. 18 (4), pp. 533-554. Chesapeake, VA: AACE.

Shechtman, N., DeBarger, A.H., Dornsife, C., Rosier, S. & Yarnall, L. (2013). Promoting Grit, Tenacity, and Perseverance: Critical Factors for Success in the 21st Century. Washington, DC: U.S. Department of Education, Department of Educational Technology.

Shute, V. J. (2007). Focus on Formative Feedback. Report RR-07-11. Princeton, NJ: Educational Testing Services, Inc.

Siegler, R. & Chen, Z. (2008). Differentiation and integration: guiding principles for analyzing cognitive change. *Developmental Science* 11(4), 433-453.

Sottilare, R. A. (2012). Considerations in the Development of an Ontology for a Generalized Intelligent Framework for Tutoring. In Proceedings of the *International Defense and Homeland Security Simulation Workshop 2012*, Vienna, Austria.

Sottilare, R. & Goldberg, B. (2012). Designing Adaptive Computer-Based Tutors to Accelerate Learning and Facilitate Retention. Journal of Cognitive Technology: Contributions of Cognitive Technology to Accelerated Learning and Expertise.

Sottilare, R., Ragusa, C., Hoffman, M. & Goldberg, B. (2013). Characterizing an adaptive tutoring learning effect chain for individual and team tutoring. In Proceedings of the Interservice/Industry Training Simulation & Education Conference, Orlando, Florida, December 2013.

Sternberg, R. (1998). Abilities are forms of developing expertise. *Educational Research,* 27 (3), pp. 11-20.

Strack, F. & Forster, J. (Eds.) (2009). *Social cognition: The Basis of Human Interaction*. London: Psychology Press.

Tangney, J. P., Baumeister, R. F. & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. Journal of Personality, 72(2), 271–322.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A. & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31,* 3-62.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A. & Wintersgill, M. (2005). The Andes physics tutoring systems: Lessons learned. *International Journal of Artificial Intelligence in Education, 15*, 147-204.

Vygotsky, L. S. (1978). *Mind in society--The development of higher psychological processes*. Cambridge: Harvard University Press.

Weschler, D. (1943). Nonintellective Factors in General Intelligence. The Journal of Abnormal and Social Psychology. 38(1), 101-103.

Wechsler, D. (1991). Wechsler Intelligence Scale for Children—Third Edition. San Antonio, TX: The Psychological Corporation.

Wolfe, P. (2010). *Brain Matters: Translating research into Classroom Practice (2nd Edition).* Alexandria, VA: Association for Supervision & Curriculum Development (ASCD)

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion, 24*(1), 3-47.

Chapter 3 – I Feel Your Pain: A Selective Review of Affect-Sensitive Instructional Strategies

Sidney D'Mello[1], Nathan Blanchard[1], Ryan Baker[2], Jaclyn Ocumpaugh[2], and Keith Brawner[3]
[1] University of Notre Dame; [2] Teachers College Columbia University; [3] U.S. Army Research Laboratory

Introduction

It is well known that students experience a range of affective states when interacting with a learning technology, be it an ITS, an educational game, a simulation environment, or even simpler interfaces that support foundational skills like reading comprehension and writing proficiency (see review in D'Mello, 2013). Positive affective states, such as contentment, delight, or pride may be triggered when a challenging problem is finally mastered. Negative affective states, such as frustration, disappointment, or anger, can occur when a learner is stuck at an impasse or in reaction to feedback from the learning environment. Learners' affect can be momentary, as in the occasional eureka moment when a major insight is obtained, prolonged as in the case of boredom for a particular topic, or dispositional when the learner is enthused or disillusioned by a particular subject across a range of lessons or even a lifetime. We also know that affect is more than a mere incidental outcome that arises during learning, but can also indirectly influence learning outcomes by modulating cognitive processes in significant ways. For example, positive affective states can inspire a broader attentional focus, which is essential for creative problem solving (Clore & Huntsinger, 2007; Isen, 2008), but can also make a learner lose focus on the task at hand. On the other hand, negative affective states can be beneficial by focusing attention (Fiedler, 2001), but can hinder problem solving by triggering a form of tunnel vision when taken to an extreme.

Affect is still a complex mystery despite almost 150 years of scientific research. Decades of research in clinical psychology have revealed that humans have a relatively poor understanding of their own affective states, including how to regulate them. In a similar vein, considerable research in interpersonal communication, social dynamics, and cultural influences has indicated that people are not very apt at accurately perceiving and responding to the affective states of others, though we overestimate our ability to do so (Kelly & Metcalfe, 2011). So what is an ITS, with impoverished sensing capabilities, a shallow understanding of its environment, and a limited action repertoire to do? Should ITSs simply proclaim affect to be an insignificant or insurmountable problem and proceed by attending to cognition as they have done in the first 20 years or so of their existence? Or should they tackle affect head-on due to its prominence and influence on cognition (and thereby learning), while at the same time being fully aware of the complexities involved in devising strategies to model affect? Our answer to the latter question is a resounding "yes," and in this chapter we discuss some affect-sensitive instructional strategies that "respond to affect." We do this by first discussing theoretical issues pertaining to affect and then by adopting a theoretical framework for the affective response strategies. The main contribution of this chapter is an exposition of six case studies, each featuring a unique affect-sensitive instructional strategy that has been developed and tested[1]. We follow this with a discussion of additional considerations for "ideal" affective strategies.

Theoretical Framework

The goal of this section is to clarify key constructs and identify an overarching theoretical framework in which to situate the affect-sensitive instructional strategies (also called affective strategies). We assume

---

[1] The reader is referred to Arroyo, Muldner, Burleson, and Woolf in chapter 7 of this volume for a discussion on additional affective strategies.

that the reader is familiar with some basics of affect science, affective computing, and ITSs, so this section is relatively brief. Although some of the claims made below are generally accepted, others are still controversial and are being actively debated in the community. We sidestep all such debates by simply asserting our working definitions and assumptions.

States, Traits, Moods, and Emotions

Let us begin by clarifying what affect is and what it is not – at least from the perspective of this chapter. Affect is a state that arises from, influences, and is influenced by neurobiology, psychophysiology, and consciousness (Izard, 2010); though, Ohman and Soares (1994) note that it can be unconsciously experienced as well. From a psychological perspective, which is the level of analysis we adopt in this chapter, an affective state is primarily a subjective feeling that influences cognition. Affect is related, but not equivalent, to motivation, attitudes, preferences, physiology, arousal, and a host of other related constructs that are often used to refer to it.

It is important to distinguish between affective traits, background moods, and emotions (Rosenberg, 1998). Affective traits are relatively stable, mostly unconscious predispositions toward particular emotional experiences. They operate by lowering the threshold for experiencing certain emotional states. As an example, a person with a hostile affective trait has a lower threshold for experiencing anger, but not necessarily other negative emotions. Moods also perform a threshold reduction function on emotional elicitation, but are considered to be more transitory and have a background influence on consciousness. Emotions are relatively brief, intense states that occupy the forefront of consciousness, have significant physiological and behavioral manifestations, and rapidly prepare the bodily systems for action. Importantly, emotions are often directed at some object (a person, an event, or even a thought), while moods are more general. These different types of affective phenomena need to be addressed differently, hence, an instructional strategy that responds to affect should be mindful of whether it is targeting a trait, a mood, or an emotion. Most of the strategies discussed here focus on *emotions,* and the term *affective state* is used to refer to both bonafide emotions (e.g., disgust, anger) as well as affect-cognitive blends like confusion and boredom. Furthermore, the chapter assumes that the management of affective traits and long-lasting moods are currently beyond the scope of a tutoring system.

Another point worth mentioning pertains to the relationship between affect and learning outcomes. It is unlikely that there are direct causal links between affect and learning. Instead, affect indirectly influences learning by modulating cognition. For example, anxiety is unlikely to directly cause poorer learning, but rather negatively influences cognition, as is the case when working memory resources are consumed by anxiety-related thoughts (e.g., fear of failure). Therefore, it is advisable for an affect-regulation strategy to consider the cognitive processes influenced by affect and alter these processes by directly changing the nature of the task or indirectly changing the underlying affect. This is the essence of an effective affective instructional strategy.

Emotion Regulation and Emotion Generation

It is useful to situate affect-sensitive instructional strategies within an overarching framework of affect. Numerous affect representation frameworks and theories exist, such as core affect (Russell, 2003), psychological construction (Barrett, 2009), basic emotions (Ekman, 1992), social perspectives (Parkinson, Fischer & Manstead, 2004), and dynamical systems models (Lewis, 2005). Although each of these can serve as viable frameworks, we choose to situate our work within the *modal model* of emotion (Gross, 2008). This model is appealing because it addresses affective strategies that are both preventative (before affect arises) as well as reactive (after affect arises).

An affective state arises when an affect-eliciting situation is experienced, attended to, and cognitively appraised. The modal model of affect assumes five broad affective regulation strategies. Four of the regulatory strategies are anticipatory, while the fifth strategy is applicable after the affect is experienced. Importantly, the processes of affect generation and affect regulation are not sequential, but demonstrate circular causality in that affect regulation can alter the affect generated, and the affect generated can trigger particular affect regulation strategies (Gross & Barrett, 2011).

The first two strategies, *situation selection* and *situation modification*, are regulatory strategies aimed at selecting or modifying contexts/situations that minimize or maximize the likelihood of experiencing certain affective states. Affect can also be regulated when a situation cannot be selected or modified via *attentional deployment*, which can involve either the avoidance of the affect-eliciting situation (distraction) or increased attention to the situation (rumination). Affect can be regulated even when a person's attention is focused on an event that has the potential to elicit a particular affective reaction. One such strategy is *cognitive change* (Dandoy & Goldstein, 1990), which involves changing the perceived meaning of a situation in order to alter its affective content. These four strategies are referred to as antecedent-focused affect regulation since they target the antecedents of affect. The fifth strategy, *response modulation,* occurs after the affective state is experienced and is referred to as response-focused affect regulation. Perhaps the most widely studied form of response modulation is *expressive suppression*, which involves a sustained effort to minimize the expression of affective behavior.

With varying levels of conscious awareness, learners continually engage in one or more of these strategies. They may select certain subjects based on perceived competence in order to alleviate anxiety (situation selection), choose topics within the selected subjects to maximize interest (situation modification), ignore states of confusion by focusing attention elsewhere or ruminate on negative feelings of frustration and despair (attentional deployment), alter attributes about failure (cognitive change), or suppress negative feelings when they arise (response modulation). An affective learning technology that operates within the processes of this framework has the following options: alter the situation (situation selection and situation modification), alter cognitions pertaining to the current situation (attentional deployment or cognitive change), or alter affective expression (response modification). The extent to which each of these strategies have been implemented and tested is discussed in the next section.

Case Studies

We now turn to six case studies to discuss affect-sensitive instructional strategies with an emphasis on systems that have been tested. It should be noted that the research on affective instructional strategies, especially those that have been systematically tested, is in its infancy. To our best knowledge, the six case studies that we review reflect much of the existing work in this area. There have been other implementations of the strategies in these case studies and these are briefly discussed as well.

Table 1 provides a loose mapping between the case studies, instructional strategies, and the five components of the modal model. We consider preventative strategies that proactively alter appraisals to prevent negative affect, as well as reactive strategies that respond to negative affect when it inevitably arises. Strategies aimed at upregulating positive affect are also discussed, though these are more infrequent. General strategies that do not explicitly target affect (e.g., edutainment) are considered to be out of scope.

Table 1. Loose mapping between affective regulation strategies and components of the modal model.

Case Study	Situation Selection	Situation Modification	Attentional Deployment	Cognitive Change	Response Modulation
Affective AutoTutor				encouraging and motivational messages	empathy and emotional displays
GazeTutor		content repetition	attentional reorientation messages		
UNC-ITSpoke		explanation-based subdialogues			
ConfusionTutor	contradictory trialogues				
Instructed Reappraisal				reappraisal	
Affective Learning Companion				affective support messages	nonverbal mirroring
Other Systems					false biofeedback

Affective AutoTutor: Empathetic, Encouraging, and Motivational Messages with Emotional Displays to Address Boredom, Confusion, and Frustration

Affective AutoTutor is a modified version of a conversational ITS that helps students develop mastery of difficult topics in Newtonian physics, computer literacy, and scientific reasoning by holding a mixed-initiative dialog in natural language (Graesser, Chipman, Haynes & Olney, 2005). The original AutoTutor system has a set of fuzzy production rules that are sensitive to the cognitive states of the learner. The Affective AutoTutor augments these rules to be sensitive to dynamic assessments of learners' affective states by addressing the presence of boredom, confusion, and frustration. The affective states are sensed by monitoring conversational cues and other discourse features, gross body movements, and facial features (D'Mello & Graesser, 2012a).

The Affective AutoTutor attempts to alter these negative states by incorporating perspectives from a number of psychological theories, including attribution theory (Weiner, 1986), cognitive disequilibrium during learning (Piaget, 1952), politeness (Brown & Levinson, 1987), and empathy (Lepper & Chabay, 1988), along with recommendations made by expert human tutors (see D'Mello et al., 2008 for details). The tutor responds with **empathetic, encouraging, and motivational dialog-moves** along with **emotional displays**. For example, the tutor might respond to mild boredom with, "This stuff can be kind of dull sometimes, so I'm gonna try and help you get through it. Let's go". A response to confusion would include attributing the source of confusion to the material: "Some of this *material* can be confusing. Just keep going and I am sure you will get it". These affective responses are accompanied by an appropriate

emotional facial expression and emotionally modulated speech (e.g., synthesized empathy or encourage-ment). These displays are considered to be a form of response modulation due to the well-established emotion contagion effect (Adolphs, 2002).

The effectiveness of the Affective AutoTutor over the original non-affective AutoTutor was tested in a between-subjects experiment where 84 learners were randomly assigned to two 30-minute learning sessions with either tutor (D'Mello et al., 2010). The results indicated that the Affective tutor helped learning for low-domain knowledge learners during the second 30-minute learning session. The Affective tutor was less effective at promoting learning for high-domain knowledge learners during the first 30-minute session. Importantly, learning gains increased from Session 1 to Session 2 with the Affective tutor whereas they plateaued with the non-affective tutor. Learners who interacted with the Affective tutor also demonstrated higher performance on subsequent transfer tests. A follow-up analysis into learners' perceptions of both tutors indicated that their perceptions of how closely the computer tutors resembled human tutors increased across learning sessions, was related to the quality of tutor feedback, and was a powerful predictor of learning (D'Mello & Graesser, 2012b). The positive change in perceptions was greater for the Affective tutor. In conclusion, this study indicated that the two affective strategies used by Affective AutoTutor, cognitive change and response modulation, improve learning, but this effect was only found for low-knowledge students.

GazeTutor: Messages to Reorienting Attention and Repetition of Unattended Content

Attentional engagement is a necessary condition for meaningful learning, so developing strategies for addressing attentional disengagement is likely to improve overall learning outcomes. Attentional disen-gagement can manifest when the learner voluntarily engages in off-task behavior (Baker, 2007) or experiences involuntary lapses in attention (mind wandering)[2]. Previous research has shown that attentional disengagement is typically a precursor to boredom (Eastwood, Frischen, Fenske & Smilek, 2012), so strategies that target it are indirectly addressing boredom. The potential effects of an attentional reengagement strategy were addressed in a study of a dialog-based learning system, called the GazeTutor. The tutor used a commercial eye tracker to monitor learners' gaze patterns in order to identify when they had attentionally disengaged (D'Mello, Olney, Williams & Hays, 2012). The tutor then attempted to re-engage learners with gaze-reorienting messages that instructed learners to pay attention to the tutor or important parts of the interface (i.e., an explanatory image). In addition, the tutor would repeat the content that was ostensibly missed due to inattention. Hence, the instructional strategy used here consisted of **direct attentional reorientation messages with content repetition.**

The efficacy of GazeTutor in promoting motivation, engagement, and learning was tested in a within-subjects experiment where 48 learners were tutored on four biology topics with both gaze-reactive and non-gaze-reactive (control condition) versions of the tutor. The results indicated that GazeTutor was successful in dynamically reorienting learners' attentional patterns to the important areas of the interface. The effectiveness of gaze-orientation faded over time but did not entirely diminish. Although gaze-reactivity did not impact self-reported motivation and engagement, post-test scores for deep reasoning questions were higher when learners interacted with the gaze-sensitive tutor. Interestingly, individual differences in scholastic aptitude moderated the impact of gaze-reactivity on learning gains. Gaze-reactivity was associated with a small improvement in overall learning for learners with average scholas-tic aptitude, but learning gains were substantially higher for learners with high aptitude and somewhat lower for their counterparts. As such, this study demonstrates that the strategies of altering the situation through content repetition and altering cognition through attentional reorientation positively affected learning, more so for learners with high scholastic aptitude.

---

[2] DeFalco, Baker, and D'Mello in chapter 4 in this volume discuss additional strategies to address disengaged behaviors.

UNC-ITSpoke: Responding to Uncertainty with Explanation-based Subdialogs

UNC-ITSPOKE is an ITS that was designed to examine whether automatic responses to learner uncertainty could improve learning outcomes (Forbes-Riley & Litman, 2007, 2009; Forbes-Riley & Litman, 2011). Uncertainty is a state that is similar to confusion and plays an important role in the process and products of learning. ITSPOKE is a speech-enabled ITS that teaches learners about various physics topics with spoken dialogs; student responses are automatically recognized with the Sphinx 2 Speech Recognizer (Litman et al., 2006). UNC-ITSPOKE extends the basic functionality of ITSPOKE with the capability to automatically detect and respond to learners' certainty/uncertainty in addition to correctness/ incorrectness of their spoken responses. Uncertainty detection is performed by extracting and analyzing the acoustic-prosodic features in learners' spoken responses in conjunction with lexical and dialog-based features.

Responses to uncertainty occurred when the student was correct in their response but uncertain about the response. This was taken to signal an impasse because the student is unsure about the state of their knowledge despite being correct. The actual response strategy involved launching **explanation-based sub-dialogs** that provided added instruction to remediate the uncertainty. This might involve additional follow-up questions (for more difficult content) or simply asserting the correct information with elaborated explanations (for easier content).

In a recent study, Forbes-Riley and Litman (2011) compared learning outcomes between 72 learners who were randomly assigned to receive adaptive responses to uncertainty (adaptive condition), no responses to uncertainty (no adapt control condition), or random responses to uncertainty (random control condition). In this later condition, the added tutorial content from the sub-dialogs was given for a random set of turns in order to control for the additional tutoring. Results indicated that the adaptive condition achieved slightly (but not significantly) higher learning outcomes than the random and control conditions. The findings revealed that it was perhaps not the presence or absence of adaptive responses to uncertainty, but the number of adaptive responses that correlated with learning performance. Unfortunately, the biggest challenge was caused by errors in automatic uncertainty detection, which reduced the number of opportunities for adaptive responses. Thus, although the findings were somewhat mixed, Forbes-Riley and Litman (2011) conclude that there is merit in offering adaptive feedback to uncertainty and that such feedback can improve learning outcomes. Further research, specifically in the area of automated uncertainty detection, is required to improve the effectiveness of an affective strategy of explanation-based sub-dialogs as a form of situation modification.

ConfusionTutor: Inducing Productive Confusion with Counterfactual and Contradictory Information

UNC-ITSpoke views uncertainty and impasses as opportunities for learning, a view that is consistent with theories that highlight the benefits of impasses (VanLehn, Siler, Murray, Yamauchi & Baggett, 2003), cognitive conflict (Limón, 2001), cognitive dissonance (Festinger, 1957), cognitive disequilibrium (Piaget, 1952), and socio-cognitive conflict (Mugny & Doise, 1978). Confusion is considered to be the affective signature of these states (D'Mello & Graesser, in press). Therefore, one hypothesis is that events that confuse learners might provide valuable learning opportunities because learners need to engage in deep cognitive activities in order to resolve their confusion. It is likely that the cognitive activities that accompany confusion resolution promote deeper learning, rather than the confusion itself.

The hypothesis that confusion can impact learning was tested by modifying an educational game, Operation ARA (Millis et al., 2011), to systematically induce confusion (D'Mello, Lehman, Pekrun & Graesser, 2014). ARA teaches scientific research methods and critical thinking skills through a series of game

modules, including those with two or more animated pedagogical agents. In the trialogues, a 3-way conversation transpired between the human student, a tutor-agent, and a student-agent. The tutor-agent was an expert on scientific inquiry, whereas the student-agent was a peer of the human learner. A series of research case studies that have a crucial experimental design flaw with respect to proper scientific methodology was presented by one of the agents. Confusion was induced by manipulating whether or not the tutor-agent and/or the student-agent provided **counterfactual** information that **contradicted** the other agent during the trialogue. The human learner was asked to intervene after each point of contradiction. If the human learner experienced uncertainty and was confused, this should be reflected in the incorrect-ness/uncertainty of his or her answer and on self-reported confusion. In some cases, the learner was presented with short instructional texts, which contained information to assist in confusion resolution.

Two experiments, with 63 and 76 learners, confirmed that contradictions increased learners' confusion. Importantly, levels of confusion moderated the impact of the contradictions on learning. Specifically, the contradictions had no effect on learning when learners were not confused by the manipulations, whereas performance on multiple-choice post-tests and on transfer tests was substantially higher when the contra-dictions were successful in confusing learners. This suggests that there are some benefits to inducing confusion if learners are *productively* instead of *hopelessly* confused. By productive confusion, we mean that the confusion is relevant to the learning content, the learner actively attends to the confusion by engaging in confusion-resolution activities, the learner has the capability to resolve the confusion, and the learning environment provides appropriate scaffolds when needed. In summary, this study showed that counterfactual and contradictory trialogues as a situation selection strategy can have significant positive impact on learning if properly directed.

Instructed Reappraisal to Increase Engagement and Positive Affect

A more recent attempt to understand emotion regulation, as defined by Gross (2008) as the physiological, behavioral, and cognitive processes that enables individuals to manage the experience and expression of emotions, is provided by Strain and D'Mello (in review). This study set out to investigate cognitive change, which involves changing the way one thinks about the situation to alter its emotional meaning. Cognitive reappraisal is suggested to be a key emotion regulation technique, yet little research in educational psychology has endeavored to understand whether cognitive change is effective during learning. Thus, the goal was to examine whether providing learners with instruction on cognitive reappraisal strategies would help them to effectively manage their emotional experiences (particularly boredom) during learning. If emotion regulation strategies are effective, then ITSs (especially those that are affect-sensitive) can encourage learners to adopt these strategies at appropriate moments.

The authors test a cognitive reappraisal strategy in the context of a 45-minute web-based self-paced learning session in which 93 participants were asked to learn about the U.S Constitution and Bill of Rights, answer simple text-based and more challenging inference questions, and report their affective states at multiple points. Participants were randomly assigned to one of three conditions: *instructed reappraisal* (IR), *error searching* (ES), or *control*. All participants were instructed that they would be reading the Constitution and Bill of Rights and answering easy and difficult questions about the material, to demonstrate that they are capable of learning a lot of information quickly and efficiently. Participants in the IR condition were asked to imagine that they were applying for a job as a copy-editor at a powerful law firm in their city. This imaginary situation involved them having to check the document for typos and grammatical errors to demonstrate their skill as copy-editors. By asking participants to imagine that they were applying for a job, it was expected that they would place more meaning on the task than if they were simply completing the task for a small payment. That is, instead of their default appraisal of reading a lengthy and boring document, they would **reappraise** the situation as being more relevant to the imagined desire to get the job. In contrast, participants in the ES condition were simply asked to perform the copy-

editing without the reappraisal component. Participants in the control condition received no special instructions about cognitive reappraisal or error searching.

Compared to the control condition, learners in the IR condition experienced more positive-activation affect (dimensionally assessed with self-reports of valence and arousal), higher engagement, lower confusion and frustration on discrete affect measures, and significantly higher learning outcomes on knowledge tests. The IR and ES conditions did not differ in arousal or engagement, but the IR condition reported significantly more positive valence, less confusion, and less frustration. The IR condition also significantly outperformed the ES condition on learning measures. This suggests the improved performance of the IR condition over the control condition was attributable to the use of the IR strategy, and not the task of error searching.

A follow-up experiment with 138 learners that compared the same IR strategy to an open-ended reappraisal (where learners adopt their own reappraisal strategy), a suppression strategy (where learners are asked to suppress all behavioral indicators of emotion), and the same control condition, found positive effects of reappraisal on positive affect, engagement, and learning (Strain & D'Mello, in review). Hence, the main conclusion is that cognitive change, even in the form of a vastly simplified reappraisal strategy used in these experiments, can be a successful method for regulating emotions and improving learning.

Affective Learning Companion with Nonverbal Mirroring and Affect Support

Burleson and Picard (2007) devised an affective strategy for an affective learning companion that helps students solve the Tower of Hanoi problem. The learning companion takes the form of an embodied conversational agent (ECA) and combines nonverbal mirroring with affective support. The **nonverbal mirroring** was accomplished by sensing learners' facial expressions, posture, electrodermal activity, and pressure exerted on the mouse. The ECA responded to this sensed data after a 4-second delay with similar facial expressions and postures, increased swaying in response to mouse pressure, and reddened skin tone to convey physiological arousal. The **affective support** intervention consisted of the ECA speaking messages that supported learners' meta-cognitive assessments of their ability to solve the problem, derived from incremental theories of intelligence (Dweck, 2006). These messages suggested that the mind is like a muscle that can be strengthened with effort.

An experiment with 61 children (11 to 13 years of age) was conducted to evaluate the affective learning companion. It employed a 2×2 between-subjects design where learners were assigned to an agent with affective support and nonverbal mirroring, task support with nonverbal mirroring, affective support with prerecorded nonverbal interaction, and task support with prerecorded nonverbal interaction. In the task support condition, the ECA provided messages pertaining to the task, but these messages did not address feelings or attempt to motivate learners. In the prerecorded nonverbal interaction condition, the ECA's nonverbal behaviors were driven by the behaviors of "average participants" from pilot studies.

The results did not yield any significant differences (main effects or interactions) on a range of outcome variables encompassing perseverance, formation of social bonds with the agent, frustration, intrinsic motivation, etc. However, exploratory follow-up-analyses did yield several interesting gender effects. For example, girls in the combined affective support plus nonverbal mirroring condition reported lower levels of frustration than girls who received each individual treatment (i.e., affective support with prerecorded nonverbal interaction or task support with nonverbal mirroring). There were additional interesting gender interactions, as discussed in Burleson and Picard (2007); however, the small sample size (roughly 7–8 per cell) warrants replication with a larger sample. The tentative results of this study appear to indicate that response modulation and cognitive change strategies can effectively be used to alter affective states, and that the learning gains induced by these strategies may be particularly effective for young girls.

Additional Implementations of Basic Strategies and Other Strategies

In addition to the six case-studies discussed in detail above, a few other studies of affective regulation strategies bear mentioning. Some systems make an inference of the underlying affective state, but do not directly attempt to detect affect. For example, Tsukahara and Ward (2001) varied the **acknowledgement** a tutor provided the student during a simple memory game by *inferring* affect based on student prosody. A small-scale user test ($N = 13$) indicated that users preferred this system compared to a control. Similarly, Andallaza and Rodrigo (2013) made inferences of student affect based on number of steps taken to solve a problem and solving duration, and responded with **motivational** messages. An experiment with 80 learners did not yield any positive effects on learning but learners indicated that they preferred the affective system compared to controls. Recently, Kelly, Heffernan, D'Mello, Namais, and Strain (2013) studied the effect of **teacher-generated motivational videos** that emphasized the value of a difficult math exercise and the importance of exerting effort toward building competence during homework completion with ASSISTments, an ITS for middle school math. They found small effects on positive valence (Experiment 1 with $N = 24$) and improved homework completion rates (Experiment 2 with $N = 60$) compared to controls, but these results warrant replication with larger samples.

There has been considerable interest in using **empathy** as an affective response strategy. This has been studied by Kim, Baylor, and Shen (2007) with 56 pre-service teachers and McQuiggan, Robison, Phillips, and Lester (2008) on 35 college students in the context of CRYSTAL ISLAND, a narrative-centered educational game. A unique feature of these studies is that the interventions were triggered from self-reports, instead of automated affect detection. Some researchers also differentiate between different types of empathetic responses (McQuiggan et al., 2008; Moridis & Economides, 2012). **Parallel empathy** simply involves mirroring the learners affective state (e.g., displaying frustration when the learner is frustrated) whereas **reactive empathy** involves performing a deeper analysis of learner affect to converge upon an appropriate response that goes beyond simple affect mirroring (e.g., displaying sadness when a learner is frustrated).

Researchers have also considered inducing states of physiological arousal in order to increase metacognitive awareness and potentially learning. Strain, Azevedo, and D'Mello (2013) used a **false biofeedback paradigm**, where learners were presented with audio stimuli of accelerated or baseline heartbeats purportedly representing their own heart beats during a challenging learning task. They found that learners self-reported experiencing more positive activating affect, made more confident metacognitive judgments, and achieved better learning when they received biofeedback compared to no biofeedback. Interestingly, these effects were only discovered for challenging questions that required inference as opposed to simpler text-based questions, and type of biofeedback (accelerated vs. baseline) had no effect.

Future Considerations

We now turn to additional issues of relevance to affect-sensitive instructional strategies, including the representation, dynamics, antecedents, and detection of affective states. Some of these aspects may be less feasible as research items in the short term given the current nascent state of the field. Nevertheless, they might serve as fruitful avenues for future research as they are likely to contribute to more "ideal" affective instructional strategies.

Affective representations can be dimensional or discrete, a topic of intense debate that has important implications for affect-sensitive instructional strategies. Valence (positive to negative) and arousal (sleepy to active) are considered to be the primary affective dimensions (Russell, 2003), though researchers have argued for additional dimensions as well (Fontaine, Scherer, Roesch & Ellsworth, 2007). Discrete affective states are usually represented as dichotomous variables (e.g., student is confused but not

frustrated, bored, anxious, etc) or ordinal variables (e.g., via Likert scales). Discrete (or categorical) representations are preferred over dimensional representations when devising affect-sensitive instructional strategies. For example, frustration and boredom are both negatively valenced, but the strategies needed to regulate the activating state of frustration are quite different than those needed for the deactivating state of boredom. However, an ITS is likely unable to differentiate between the two states using only valance and arousal. For this reason, discrete representations are better able to inform affective instructional strategies.

Affective dynamics, in the form of timing and intensity, are of singular importance. Some affective states are ephemeral (e.g., surprise, eureka moments), while others are more persistent (e.g., boredom, anxiety) (Baker, D'Mello, Rodrigo & Graesser, 2010; D'Mello & Graesser, 2011). A state can also exhibit ephemeral properties in some situations while demonstrating persistence in others; these differences in temporal duration can differentially impact learning. For example, experiences of confusion that are immediately resolved are expected to have little to no effect on learning, whereas persistent confusion that is never resolved might be negatively related to learning (D'Mello & Graesser, in press). Timing and intensity of affect can also interact in striking ways. A long-lasting but low-intensity state of anxiety might not be very impactful, but a single episode of intense embarrassment or anger can have long-lasting negative consequences (e.g., dislike for an ITS based in one unpleasant interaction can engender negative feelings toward an entire course). Hence, it is advisable for an affect-sensitive instructional strategy to be sensitive to the timing and intensity of affect.

Affect-inducing events have a singular effect on the affective states generated and how they are expressed. Thus, successfully regulating an affective state entails understanding the affect-inducing event and the appraisals of the event that gave rise to the state. Boredom offers a convenient example. According to Pekrun's *control-value* theory of academic emotion, subjective appraisals of control and value of a learning activity are critical predictors of boredom and other academic emotions (Pekrun, 2010). Subjective control pertains to the perceived influence that a learner has over the activity and its outcomes, while subjective value represents the perceived value of the activity. Boredom is expected to be heightened when learners perceive low value in the outcome of the activity, and both when control is too low (challenge exceeds skill) or too high (skill exceeds challenge). An intervention that attempts to reengage bored learners by emphasizing the value of the learning activity will miss its mark entirely when the underlying cause of boredom is due to a lack of control. It can even have negative consequences, as noted by Durik and Harackiewicz (2007) who found that informing low-competence students (low control) about the relevance of math material for their lives (value manipulation) actually undermined value because it was perceived as threatening. The important message here is that an effective affect-sensitive instructional strategy should be sensitive to the antecedents of the affective state in addition to the affective state itself.

Affect detection is usually a first step for affect-sensitive instructional strategies. Affect detection is perhaps the most actively explored subfield of affective computing (see reviews by Calvo & D'Mello, 2010; D'Mello & Kory, 2012; Zeng, Pantic, Roisman & Huang, 2009), but like much of the affective sciences is inherently imperfect and is unlikely to ever reach perfection. How can we tailor instructional strategies in anticipation of imperfect affect detection? In addition, we outlined additional considerations for affective instructional strategies in this section. We advocated a focus on discrete affect representations, an emphasis on the timing and intensity of affective states, and on considering the antecedents of affect while tailoring instructional strategies. These pose additional challenges for affect detectors that are now faced with the task of detecting intensity, duration, and antecedents, in addition to the already challenging task of basic affect detection. Therefore, progress in affect detection is essential before some of these "ideal" affect-sensitive instructional strategies can be effective.

Conclusions

ITSs have been devised to provide more fine-grained domain and student modeling, allowing instruction to be tailored in a more highly individualized manner than their computer-based learning predecessors (Psotka, Massey & Mutter, 1988). Their effectiveness compared to other forms of instruction is impressive as documented in recent reviews and meta-analyses (Steenbergen-Hu & Cooper, in press; VanLehn, 2011), but this positive news has been tempered by the suggestion that improvements in the effectiveness of ITSs have somewhat leveled off, reaching what VanLehn (2008) refers to as the *interaction plateau*. Might this plateau be partially attributed to the fact that ITSs have traditionally focused on modeling cognition while largely ignored affect and motivation? If so, there might be the added benefits to improving ITS effectiveness by devising strategies to respond to these non-cognitive aspects of learning. Here, we considered the possibility of increasing the bandwidth of ITS adaptivity by modeling student affect.

This chapter described case studies of six systems that implemented 12 affect-sensitive instructional strategies: encouragement, motivational messages, empathy, emotional displays, attentional reorientation messages, content repetition, explanation-based subdialogs, contradictory trialogues, instructed reappraisal, affective support messages, nonverbal mirroring, and false biofeedback. These strategies are impressive in breadth as they cover cognitive, affective, motivational, nonverbal, and metacognitive aspects of learning. Systems that have implemented these strategies have had some success in terms of promoting positive outcomes like engagement, persistence, and learning. Although there was considerable variability in effectiveness of the affective strategies, one consistent finding is that effectiveness almost always varied as a function of differences in individual attributes (e.g., gender, prior knowledge, scholastic aptitude) and/or aspects of the learning session (e.g., content difficulty, outcome measure). This suggests that there are limits to the current one-size-fits-all approach, where variants of the same strategy are indiscriminately used for all learners and in all situations. The strategies need to be more focused by configuring them to be sensitive to learner attributes, nuances of the learning session (affect-eliciting events), and different manifestations of the same affective state (e.g., different types of boredom). This level of adaptivity will require continual improvements in automated affect sensing and context modeling, coupled with a deeper understanding of affect during learning. We consider this to be the next grand challenge for the field of affect-sensitive learning environments.

Acknowledgment

References

Adolphs, R. A. (2002). Neural systems for recognizing emotions. *Current Opinion in Neurobiology, 12*, 169-177.

Andallaza, T. C. S. & Rodrigo, M. M. T. (2013). Development of an Affect-Sensitive Agent for Aplusix. In H. C. Lane, K. Yacef, J. Mostow & P. Pavlik (Eds.), *Proceedings of the 16th conference on Artificial Intelligence in Education* (pp. 575-578). Springer: Berlin/Heidelberg.

Arroyo, I., Muldner, K., Burleson, W. & Woolf, B. (in press). Design Recommendations for Adaptive Intelligent Tutoring Systems: Adaptive Instructional Strategies (Volume 2). In R. Sottilare, A. Graesser, X. Hu & B. Goldberg (Eds.), *Design Recommendations for Adaptive Intelligent Tutoring Systems: Adaptive Instructional Strategies (Volume 2)*: U.S. Army Research Laboratory.

Baker, R. (2007). Modeling and Understanding Students' Off-Task Behavior in Intelligent Tutoring Systems *Conference on Human Factors in Computing Systems, Vol 1 and 2* (pp. 1059-1068).

Baker, R., D'Mello, S., Rodrigo, M. & Graesser, A. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments *International Journal of Human-Computer Studies, 68*(4), 223-241.

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. [Review]. *Cognition & Emotion, 23*(7), 1284-1306. doi: 10.1080/02699930902985894

Brown, P. & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Burleson, W. & Picard, R. (2007). Evidence for gender specific approaches to the development of emotionally intelligent learning companions. *IEEE Intelligent Systems, 22*(4), 62-69.

Calvo, R. A. & D'Mello, S. K. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1*(1), 18-37. doi: 10.1109/T-AFFC.2010.1

Clore, G. L. & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences, 11*(9), 393-399. doi: 10.1016/j.tics.2007.08.005

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105*(4), 1082-1099.

D'Mello, S. & Graesser, A. (2012a). AutoTutor and Affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems, 2*(4), 23:22-23:39.

D'Mello, S. & Graesser, A. (2012b). Malleability of students' perceptions of an affect-sensitive tutor and its influence on learning. In G. Youngblood & P. Mccarthy (Eds.), *Proceedings of 25th Florida Artificial Intelligence Research Society Conference* (pp. 432-437). Menlo Park, CA: AAAI Press.

D'Mello, S. & Graesser, A. (in press). Confusion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *Handbook of Emotions and Education*: Taylor & Francis.

D'Mello, S., Jackson, G., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., El Kaliouby, R., Picard, R. & Graesser, A. (2008). *AutoTutor detects and responds to learners affective and cognitive states.* Paper presented at the Proceedings of the Workshop on Emotional and Cognitive Issues in ITS held in conjunction with the Ninth International Conference on Intelligent Tutoring Systems, Montreal, Canada.

D'Mello, S. & Kory, J. (2012). Consistent but Modest: Comparing multimodal and unimodal affect detection accuracies from 30 studies. In L.-P. Morency, D. Bohus, H. Aghajan, A. Nijholt, J. Cassell & J. Epps (Eds.), *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (pp. 31-38). New York: ACM

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Perkins, L. & Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In J. Kay & V. Aleven (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems* (pp. 245-254). Berlin / Heidelberg: Springer.

D'Mello, S., Lehman, S., Pekrun, R. & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*(1), 153-170.

D'Mello, S., Olney, A., Williams, C. & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies, 70*(5), 377-398.

D'Mello, S. & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion, 25*(7), 1299-1308.

Dandoy, A. C. & Goldstein, A. G. (1990). The Use of Cognitive Appraisal to Reduce Stress Reactions - a Replication. *Journal of Social Behavior and Personality, 5*(4), 275-285.

DeFalco, J., Baker, R. & D'Mello, S. (in press). Addressing Behavioral Disengagement in Online Learning. In R. Sottilare, A. Graesser, X. Hu & B. Goldberg (Eds.), *Design Recommendations for Adaptive Intelligent Tutoring Systems: Adaptive Instructional Strategies (Volume 2)*: Army Research Laboratory.

Durik, A. M. & Harackiewicz, J. M. (2007). Different strokes for different folks: How individual interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology, 99*(3), 596-610.

Dweck, C. (2006). *Mindset*. New York: Random House.

Eastwood, J. D., Frischen, A., Fenske, M. J. & Smilek, D. (2012). The Unengaged Mind Defining Boredom in Terms of Attention. *Perspectives on Psychological Science, 7*(5), 482-495.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3-4), 169-200.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

Fiedler, K. (2001). Affective states trigger processes of assimilation and accommodation. *Theories of mood and cognition: A user's guidebook*, 85-98.

Fontaine, J., Scherer, K., Roesch, E. & Ellsworth, P. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18*(12).

Forbes-Riley, K. & Litman, D. (2007). *Investigating Human Tutor Responses to Student Uncertainty for Adaptive System Development*. Paper presented at the Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction, Lisbon, Portugal.

Forbes-Riley, K. & Litman, D. (2009). Adapting to student uncertainty improves tutoring dialogues. In V. Dimitrova, R. Mizoguchi & B. Du Boulay (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 33-40). Amsterdam: IOS Press.

Forbes-Riley, K. & Litman, D. J. (2011). Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication, 53*(9-10), 1115-1136. doi: 10.1016/j.specom.2011.02.006

Graesser, A., Chipman, P., Haynes, B. & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education, 48*(4), 612-618. doi: 10.1109/TE.2005.856149

Gross, J. (2008). Emotion regulation. In M. Lewis, J. Haviland-Jones & L. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 497-512). New York, NY: Guilford.

Gross, J. J. & Barrett, L. F. (2011). Emotion generation and emotion regulation: One or two depends on your point of view. *Emotion Review, 3*(1), 8-16.

Isen, A. (2008). Some ways in which positive affect influences decision making and problem solving. In M. Lewis, J. Haviland-Jones & L. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 548-573). New York, NY: Guilford.

Izard, C. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review, 2*(4), 363-370. doi: 10.1177/1754073910374661

Kelly, K., Heffernan, N., D'Mello, S., Namais, J. & Strain, A. (2013). Adding Teacher-Created Motivational Video to an ITS. In C. Boonthum-Denecke & G. M. Youngblood (Eds.), *Proceedings of 26th Florida Artificial Intelligence Research Society Conference* (pp. 503-508). Menlo Park, CA: AAAI Press.

Kelly, K. & Metcalfe, J. (2011). Metacognition of emotional face recognition. *Emotion, 11*(4), 896-906.

Kim, Y., Baylor, A. & Shen, E. (2007). Pedagogical agents as learning companions: the impact of agent emotion and gender. *Journal of Computer Assisted Learning, 23*(3), 220-234.

Lepper, M. & Chabay, R. (1988). Socializing the intelligent tutor: Bringing empathy to computer tutors. In H. Mandl & A. Lesgold (Eds.), *Learning Issues for Intelligent Tutoring Systems* (pp. 242-257). Hillsdale, NJ: Erlbaum.

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and Brain Sciences, 28*(2), 169-245.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction, 11*(4–5), 357-380. doi: 10.1016/s0959-4752(00)00037-2

Litman, D., Rose, C., Forbes-Riley, K., VanLehn, K., Bhembe, D. & Silliman, S. (2006). Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence In Education, 16*(2), 145-170.

McQuiggan, S. W., Robison, J. L., Phillips, R. & Lester, J. C. (2008). *Modeling parallel and reactive empathy in virtual agents: An inductive approach.* Paper presented at the Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1.

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A. & Halpern, D. (2011). Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou & J. Lakhmi (Eds.), *Serious games and edutainment applications* (pp. 169-196). London, UK: Springer-Verlag.

Moridis, C. & Economides, A. (2012). Affective Learning: Empathetic Agents with Emotional Facial and Tone of Voice Expressions. *IEEE Transactions on Affective Computing, 3*(3), 260-272.

Mugny, G. & Doise, W. (1978). Socio-cognitive conflict and structure of individual and collective performances. *European Journal of Social Psychology, 8*(2), 181-192.

Ohman, A. & Soares, J. (1994). Unconscious Anxiety - Phobic Responses to Masked Stimuli. *Journal of Abnormal Psychology, 103*(2), 231-240.

Parkinson, B., Fischer, A. H. & Manstead, A. S. (2004). *Emotion in social relations: Cultural, group, and interpersonal processes*: Psychology Press.

Pekrun, R. (2010). Academic emotions. In T. Urdan (Ed.), *APA educational psychology handbook* (Vol. 2). Washington, DC: American Psychological Association.

Piaget, J. (1952). *The origins of intelligence*. New York: International University Press.

Psotka, J., Massey, D. & Mutter, S. (1988). *Intelligent tutoring systems: Lessons learned*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosenberg, E. (1998). Levels of analysis and the organization of affect. *Review of General Psychology, 2*(3), 247-270. doi: 10.1037//1089-2680.2.3.247

Russell, J. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172.

Steenbergen-Hu, S. & Cooper, H. (in press). A Meta-Analysis of the Effectiveness of Intelligent Tutoring Systems on College Students' Academic Learning. *Journal of Educational Psychology*.

Strain, A., Azevedo, R. & D'Mello, S. K. (2013). Using a false biofeedback methodology to explore relationships between learners' affect, metacognition, and performance. *Contemporary Educational Psychology, 38*(1), 22-39.

Strain, A. & D'Mello, S. (in review). Emotion regulation during learning: The enhancing effect of cognitive reappraisal.

Tsukahara, W. & Ward, N. (2001). Responding to subtle, fleeting changes in the user's internal state *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 77-84). New York, NY: ACM.

VanLehn, K. (2008). *The Interaction Plateau: Answer-based Tutoring < Step-based Tutoring = natural tutoring.* Paper presented at the Intelligent Tutoring Systems.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T. & Baggett, W. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209-249. doi: 10.1207/S1532690XCI2103_01

Weiner, B. (1986). *An attributional theory of motivation and emotion*. New York: Springer-Verlag.

Zeng, Z., Pantic, M., Roisman, G. & Huang, T. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(1), 39-58.

Chapter 4 – Addressing Behavioral Disengagement in Online Learning

Jeanine A. DeFalco[1], Ryan S. Baker[1], and Sidney K. D'Mello[2]
[1]Teachers College, Columbia University; [2]University of Notre Dame

Disengaged Behavior – A Problem in Online Learning

In recent years, there has been increasing awareness that behavioral disengagement plays an important role in online learning. Not only are some forms of behavioral disengagement associated with lower learning gains in the short term (in the case of online learning, see Gobel, 2008; Cocea et al., 2009), behavioral disengagement is also associated with lower long-term academic performance (Finn & Owings, 2006; Wang & Eccles, 2012; Pardos et al., 2013) and even whether learners advance in their academic career years later (Ensminger & Slusarcick, 1992; San Pedro et al., 2013).

Correspondingly, there has been increasing interest in developing interventions that address learners' behavioral disengagement, reducing it and/or mitigating its effects on learning and long-term academic achievement. In this chapter, we discuss several types of interventions, and potentially fruitful directions for the next generation of adaptive interventions, to reduce behavioral disengagement, discussing how these interventions can be incorporated into the GIFT framework for broad dissemination.

Within this chapter, we conceptualize behavioral engagement (and disengagement) within the framework provided by Fredericks, Blumenfeld, and Paris (2004). They define school engagement in terms of three components: behavioral engagement, emotional/affective engagement, and cognitive engagement. Within this chapter, we focus on behavioral engagement (the other types of engagement are discussed in separate chapters in this volume). Behavioral engagement is defined by Fredericks and colleagues (2004) as participation, effort, persistence, and positive conduct while directly involved in a set of activities: "Behavioral engagement is most commonly defined in three ways. The first definition entails positive conduct, such as following the rules and adhering to classroom norms, as well as the absence of disruptive behaviors such as skipping school and getting in trouble […] The second definition concerns involvement in learning and academic tasks and includes behaviors such as effort, persistence, concentration, attention, asking questions, and contributing to class discussion […]. A third definition involves participation in school-related activities such as athletics or school governance" (Fredricks, Blumenfeld & Paris, 2004, p. 62).

We define behavioral disengagement in terms of the first definition, where students fail to follow the rules or expectations for the activity, engaging instead in behaviors outside of the norms or expectations, such as ceasing to participate in the activity or participating in it in an undesired and inappropriate fashion.

One of the core types of disengaged behavior, seen across a wide range of interactive learning environments, is gaming the system (Baker, Corbett, Koedinger & Wagner, 2004). Gaming the system is defined as systemically taking advantage of a software's help and feedback feature to advance through the tutoring curriculum while bypassing actively thinking about the learning material (Baker et al., 2004). Examples include systematic guessing and clicking through hints to obtain answers, but different gaming behaviors such as intentionally making spam posts and making spam responses to those spam posts are seen in other learning environments (Cheng & Vassileva, 2005). Among disengaged behaviors, gaming the system has been found to be particularly strongly associated with learner outcomes, including short-term learning (Cocea et al., 2009), longer-term learning outcomes (Pardos et al., 2013), and college attendance (San Pedro et al., 2013).

In addition to gaming the system, a range of other disengaged behaviors are seen in online learning environments. For example, learners can go completely off-task (Karweit & Slavin, 1982), ceasing to participate in the learning task. Off-task behavior's relationship to learning is typically negative, but not strongly so (Goodman et al., 1990) – and it may serve as a way of disrupting boredom, which is more strongly associated with negative learning outcomes (Baker, Moore et al., 2011). Indeed, research has shown that off-task behavior during expert tutoring sessions can improve motivation, build rapport between the tutor and learner, and allow for periodic rest (Lehman, Cade & Olney, 2010). In online learning, there have been multiple studies finding no relationship between off-task behavior and learning or other outcomes (Cocea et al., 2009; Pardos et al., 2013; San Pedro et al., 2013); the reasons for this are not yet known.

Some learners exhibit behaviors within the learning environment that are unrelated to the learning task – this behavior, sometimes called off-task behavior (Rowe et al., 2009) and sometimes called WTF behavior ("without thinking fastidiously" – Wixon et al., 2012), can manifest in many ways. For example, in a multi-user virtual environment, learners may obtain virtual cacti and place them in on a virtual patient, or climb virtual buildings (Sabourin, Rowe, Mott & and Lester, 2013). In a simulation microworld, learners may engage variables in rapid succession or pause and un-pause a simulation very quickly and repeatedly (Wixon et al., 2012). In one report, no relationship was found between this behavior and learning (Rowe et al., 2009), but its relationship to learning has not been studied in other learning environments.

Learners can also make careless errors, an error that a student makes when answering a question that they know how to do with no obvious reason why they erred (Clements, 1982). Careless errors are seen both in offline learning and assessment (e.g., Clements, 1982), and in online learning (San Pedro, Baker & Rodrigo, 2011). Careless errors are typically a behavior characteristic of generally more successful learners (Clements, 1982), but are still associated with negative outcomes after learner knowledge is controlled for (Baker et al., 2010; San Pedro et al., 2013).

Though these are the most studied disengaged behaviors in the context of online learning, other behaviors have also been seen, such as killing your teammates in military simulations for no apparent reason (Sottilare, 2013).

Addressing Gaming the System in Online Learning

Given the relatively strong evidence that gaming the system is associated with worse outcomes for learners, it is perhaps unsurprising that it has been a particular focus of research to address disengaged behaviors in online learning. There have been many approaches to addressing gaming in online learning, including attempting to make gaming more difficult, detecting gaming and employing embodied agents that look unhappy when students game, changing the incentive structure to de-incentivize gaming, giving meta-cognitive messages about how to learn effectively, and using visualizations of the student's behavior to show them how much they have been gaming.

There are several ways to make gaming more difficult. The most popular strategy employed to accomplish this goal is introducing delays to each level of on-demand hints (clicking rapidly through on-demand hints is one of the most popular ways for learners to game the system). With delayed hints, each time a learner receives a hint, there is a pre-determined amount of time they must wait before they can request another hint (Murray & VanLehn, 2005; Beck, 2005). However, this approach has thus far been ineffective because learners find alternative ways to game the system. In addition, it has the drawback that it discourages some appropriate types of hint use.

Both emotional expressions (on the part of an embodied agent) and changing the incentive structure to de-incentivize gaming were incorporated into Scooter the Tutor (Baker et al., 2006). Scooter the Tutor was an embodied agent that responded when a learner's behavior indicated that they were gaming the system (according to an automated detector of gaming – cf. Baker et al., 2008). Scooter responded by looking unhappy when the learner gamed (and telling the student not to game), and if the gaming behavior persisted, Scooter gave supplementary exercises that slowed the learner down (while also giving the learner an alternate way to learn material bypassed by gaming). In studies in the United States, Scooter reduced gaming and improved learning (Baker et al., 2006; Belmontes et al., 2011), with the supplementary exercises having more effect than the emotional expressions. However, learners disliked Scooter (Rodrigo et al., 2012). In the Philippines, Scooter actually increased the amount of apparent gaming, as learners appreciated Scooter's supplementary exercises and intentionally clicked through hints in order to obtain them (Rodrigo et al., 2012).

A third approach, providing meta-cognitive messages on how to learn more effectively, was adopted by Roll and colleagues (2007). The Help Tutor system responds to gaming the system behavior by giving meta-cognitive feedback, suggesting students should request a hint or slow down and read hints more carefully – for example, "It may not seem like a big deal, but hurrying through these steps may lead to later errors. Try to slow down." (Roll et al., 2007, p. 205). Although this system reduced gaming behaviors, it did not have a positive impact on learning (Roll et al., 2007).

Another approach, visualizing gaming behavior, was combined with text messages (Walonoski & Heffernan, 2006). In this work, a knowledge-engineered gaming detection model was used to select when students would receive interventions. When a learner was assessed to be gaming, the learner received text messages that asked (for example) whether the learner was guessing or actually needed the hint requested. In addition, the screen continually displayed a graphical visualization of learner actions and progress, which displayed gaming behavior as well as other student actions, in a way that was visible to both the student and the teacher. This combined intervention of dynamic active (text messages) and dynamic passive interventions (the visualization) resulted in reduced gaming during the intervention, (Walonoski & Heffernan, 2006). This intervention's effects on domain learning outcomes have not yet been studied.

Another category of gaming intervention is visualizations between problems. In Arroyo et al. (2007), how much the student had gamed the system was visualized between problems, in combination with detailed messages about appropriate meta-cognitive behavior encouraging students to slow down and attentively read problems and hints , e.g., "Dear Ivon, We think this will make you improve even more: Read the problem thoroughly. If the problem is just too hard, then ask for a hint. Read the hints CAREFULLY. When a hint introduces something that you didn't know, write it down on paper for the next time you need it" (Arroyo et al., 2007, p. 2). The system also included messages that encouraged students to think about the problem and guess at the solution, and ask for hints if the guess was wrong, e.g., "Dear Ivon, Think through the problem thoroughly and make a guess. If your guess is wrong, no problem, just ask for a hint. If you need more hints, keep clicking on help" (Arroyo et al., 2007, pg. 2).

Arroyo and colleagues (2007) argued that giving feedback on gaming between problems could improve behavior and learning without disrupting problem-solving activity, in addition to increasing the chances of immediate reengagement after seeing an intervention. When evaluated, this intervention led to a lower degree of gaming the system (Arroyo et al., 2007). The between-problem visualizations of how much the student gamed also led learners to spend more time on the subsequent problem. The combined intervention was associated with improved learning of domain content, as well as improving learner attitudes toward the system – a strong contrast to the negative attitudes of students toward Scooter the Tutor.

In a follow-up study, between-problem visualizations were not given, but three types of intervention messages were used: attribution interventions, effort-affirmation interventions, and strategic interventions

(Arroyo et al., 2010). Attribution interventions messages were given when a student faced a new problem, e.g., "I found out that people have myths about math, think that only some people are good in math. Truth is we can all be good in math if we try," (Arroyo et al., 2010, p. 5). Effort-affirmation intervention messages were generated when a learner achieves a correct answer; different messages were given depending on whether a correct answer was generated with effort or no effort; for effort: "Keep in mind that when we are struggling with a new skill we are learning and becoming smarter!"; for no effort: "We will learn new skills only if we are persistent. If we are very stuck, let's call the teacher or ask for a hint from Wayang!" (Arroyo et al., 2010, p. 5). Finally, strategic interventions focused on meta-cognitive strategies that could be used when a student was correct or incorrect; for correct: "We are making progress. Can you think of what we have learned in the last 5 problems?"; for incorrect: "Are we using a correct strategy to solve this? What are the different steps we have to carry out to solve this one?" (Arroyo et al., 2010, p. 5). This system resulted in less gaming the system, less frustration, and more interest as compared to a control condition. However, there was no impact on learning.

A similar result was found by Verginis et al (2011), who incorporated indicators of recent student gaming and other behaviors in a screen separate from the problem (cf. Arroyo et al., 2007), as well as providing comparisons of how much the student engaged in these behaviors compared to other students. Their article found that 39 of 73 students who were initially engaging in gaming behaviors ceased to engage in those behaviors over the course of using their system, a proportion that is not significantly different than chance according to a sign test. They did find that students who reduced their disengaged behavior achieved significantly better learning than students who did not reduce their disengaged behavior.

Across these papers, it is clear that there are several methods that can effectively reduce gaming the system. However, the only two methods that have been shown to both reduce gaming and improve learning are the types of supplementary exercises given in Scooter the Tutor, and the combination of between-problem visualizations and meta-cognitive messages. Further work may better elucidate the benefits of these approaches, and of other approaches.

Addressing Other Disengaged Behaviors In Online Learning

Thus far, there has been considerably less work addressing disengaged behaviors beyond gaming the system in online learning systems. One of the few examples of this work is seen in Hughes (2010), which proposed using Scooter the Tutor for off-task interventions as well as for gaming interventions. Specifically, if a student was off-task according to the off-task detector (Baker, 2007), then the screen would go black and a pop-up would appear with Scooter asking if the student is still at their workstation. The idea behind this intervention is that it would both encourage the student to return to work and would also attract the attention of a teacher to pay attention to the absent learner (Hughes, 2010). However, these designs were not implemented or tested in a running system.

Interventions for disengaged behaviors other than gaming the system are much more common outside the context of interactive and online learning. Some of these interventions, and the communities producing them, are discussed in the following section on future directions.

Future Directions

In this chapter, we have discussed methods that designers of interactive learning environments have used to remediate or otherwise address disengaged behaviors, particularly gaming the system. Some of these efforts have been quite successful, such as providing visualizations of disengaged behaviors between problems (Arroyo et al., 2007). However, this area of research has not scaled as of the time of this

writing. Part of the reason for this is that these interventions are time-consuming to implement, and existing ITS infrastructures are typically not designed with these types of interventions in mind. This is an excellent opportunity for a framework such as GIFT. By explicitly incorporating infrastructure-level support for developers to create these types of interventions (messages, visualizations, and embodied agents), and link them to automated detectors, it will become much easier to develop and test these types of interventions.

An additional important future direction comes from the areas of expertise brought to bear on the design of these interventions. The methods of the positive behavior support (PBS) and behavior modification (Weiss et al., 2009) used in communities of research and practice are particularly relevant to this type of intervention. These communities have been working to develop classroom practices that reduce and remediate off-task and other disengaged behaviors (termed problem behaviors in these communities) for decades (Weiss et al., 2009). And yet, there has been almost no cross-fertilization between these communities; with the exception of the participation of one behavior modification researcher in the design of Scooter the Tutor, none of the approaches discussed above involved participation from researchers or practitioners in these communities.

PBS includes integrating academics, instruction, and achievement with strategies to reinforce discipline, student self-management, and behavior management to promote cooperative and academically engaged learners (Weiss et al., 2009; Knoff, 2012). PBS entails specifying expected behaviors, teaching these expectations to learners, recognizing behavior that meet these expectations, remediating behavior that does not meet expectations through imposed consequences, and monitoring and analyzing the implementation of PBS to adjust future PBS strategies (McKevitt et al., 2012).

Many approaches and findings from these communities have relevance to the problem of reducing disengaged behaviors. For example, Kraemer and colleagues (2012) review two classroom-wide PBS interventions entitled "The Mystery Motivator" and "Get 'Em On Task*,"* (Kraemer et al., 2012, p. 163). In The Mystery Motivator, students are rewarded for engaging in positive behaviors selected by a teacher or other adult, such as staying in one's seat or working quietly, and receive a reward from a box corresponding to the day the target behavior is achieved. If the box has a Mystery Motivator symbol, a learner can chose a reward from a special reward menu.

In the Get 'Em On Task intervention, a computer program generates randomly timed sounds for monitoring student behavior (Kraemer et al., 2012). A teacher can use a classroom computer to generate random signals from 0 to 100 to sound on the hour as well as program additional random signals throughout the day (Kraemer et al., 2012). When these sounds occur, the teacher assigns points to learners who are on task, students who are off-task receive no points, (Kraemer et al., 2012), and points can subsequently be exchanged for rewards. The effects of the interventions indicated that while both the Mystery Motivator and the Get 'Em On Task interventions were effective in decreasing off-task behavior in comparison to no intervention, Get 'Em On Task had a difference in decrease of overall off-task behavior of 16.75% as compared to Mystery Motivator, (Kraemer et al., 2012).

Another approach, the *response to intervention* model (National Center on Response to Instruction, 2010), integrates PBSs into learning experiences. In this approach, the first line of intervention includes *surface management techniques* for behavior management (Sayesk & Brown, 2011). Surface management techniques include the following: (1) ignoring attention-seeking behavior; (2) signal interference, nonverbal signals such as a sound or flicker of lights, to remind learners about rules; (3) proximity and touch control, presence of a teacher nearby; (4) directly addressing a learner by name when their attention is wandering; (5) deliberate, sincere attention by instructor demonstrating concern for learner; (6) tension decontamination through humor; (7) hurdle help, providing instructional support in place of a reprimand; (8) interpretation as interference, helping students understand a confusing or frustrating experience;

(9) regrouping, physically reconfiguring a space in the classroom; (10) restructuring, changing an activity to stem off disruptive behavior; (11) direct appeal, reminder of rules; (12) limitation of space and tools, limiting learners access to materials that might tempt problem or disengaged behaviors; (13) removing a learner from the classroom to complete a neutral task without the negative connotation of being thrown out of a class; (14) permission and authoritative verboten ("No!"), clearly and succinctly communicating a particular behavior is not tolerated; and (15) promises and rewards, delivered randomly and at unexpected times, (Sayesk & Brown, 2011). While some of these approaches may not be immediately feasible in online learning, many could be realized by a pedagogical or non-player character (NPC) agent in some form.

By bringing in the ideas and successful approaches from other communities, we may be able to find better ways to address disengaged behaviors, guiding learners to engage in appropriate behaviors and helping them to learn more effectively as a result. By embedding support for creating effective interventions into architectures like GIFT, we may be able to realize these interventions at scale, creating significant positive impact on learners.

Acknowledgments

References

Aleven, V. (2001) Helping Students to Become Better Help Seekers: Towards Supporting Metacognition in a Cognitive Tutor. Paper presented at *German-USA Early Career Research Exchange Program: Research on Learning Technologies and Technology-Supported Education.* Tubingen, Germany.

Aleven, V., McLaren, B., Roll, I. & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence and Education*. 16, 101-128.

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D., Barto, A., Mahadevan, S. & Woolf, B.P. (2007) Repairing Disengagement with Non-Invasive Interventions. *Proceedings of the 13th International Conference of Artificial Intelligence in Education.* IOS Press.

Arroyo, I., Woolf, B., Royer, J., Tai, M., Muldner, K., W., Cooper, D. (2010). Gender Matters: The Impact of Animated Agents on Students' Affect, Behavior and Learning. Technical report UM-CS-2010-020. Computer Science Department, UMASS Amherst.

Babin, L., Tricot, A. & Mariné, C. (2008). Seeking and providing assistance while learning to use information systems. *Computers & Education*. 53, 1029-1039.

Baker, R.S.J.d., Corbett, A.T., Gowda, S.M., Wagner, A.Z., MacLaren, B.M., Kauffman, L.R., Mitchell, A.P., Giguere, S. (2010). Contextual Slip and Prediction of Student Performance After Use of an Intelligent Tutor. *Proceedings of the 18th Annual Conference on User Modeling, Adaptation, and Personalization*. 52-63.

Baker, R., Corbett, A., Koedinger, K., Evenson, S., Roll, I., Wagner, A., Naim, M., Raspat, J., Baker, D. & Beck, J. (2006). Adapting to when student game an intelligent tutoring system. *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*. 392-401.

Baker, R., Corbett, A.T., Koedinger, K.R., Wagner, A.Z. (2004). Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System." *Proceedings of ACM CHI 2004: Computer-Human Interaction*. 383-390.

Baker, R.S.J.d., Gowda, S.M., Corbett, A.T. (2011). Automatically Detecting a Student's Preparation for Future Learning: Help Use is Key. *Proceedings of the 4th International Conference on Educational Data Mining.* 179-188.

Baker, R.S.J.d., Mitrovic, A., Mathews, M. (2010). Detecting Gaming the System in Constraint-Based Tutors. *Proceedings of the 18th Annual Conference on User Modeling, Adaptation, and Personalization.* 267-278.

Baker, R.S.J.d., Pardos, Z., Gowda, S., Nooraei, B., Heffernan, N. (2011). Ensembling Predictions of Student Knowledge within Intelligent Tutoring Systems. *Proceedings of 19th International Conference on User Modeling, Adaptation, and Personalization.* 13-24.

Baker, R.S.J.d., Rossi, L.M. (2013). Assessing the Disengaged Behavior of Learners. In Sottilare, R., Graesser, A., Hu, X. & Holden, H. (Eds.) *Design Recommendations for Intelligent Tutoring Systems -- Volume 1 -- Learner Modeling.* U.S. Army Research Lab, Orlando, FL, 155-166.

Beck, J. (2005). Engagement tracing: using response times to model student disengagement. *Proceedings of the 12th International Conference on Artificial Intelligence in Education.* Amsterdam. 88-95.

Belmontes, R., Fontaine, S. & Genato, R. (2011). Gaming in middle schools: An interactive qualifying project report. Unpublished masters thesis. Worcester Polytechnic Institute, Worcester, MA.

Betoret, F. & Artiga, A. (2011). The relationship among student basic need satisfaction: Approaches to learning, reporting of avoidance strategies and achievement. *Electronic Journal of Research in Educational Psychology. 9*(2), 463-496.

Clements, K. (1982). Careless errors made by sixth-grade children on written mathematical tasks. *Journal for Research in Mathematics Education.* 13, 136-144.

Ebner, M. & Holzinger, A. (2007). Successful implementation of user-centered game based learning in higher education: An example from civil engineering. *Computers & Education. 49*(3), 873-890.

Ensminger, M. E. & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education.* 65, 95-113.

Finn, J. D. & Owings, J. (2006). The adult lives of at-risk students: The roles of attainment and engagement in high school, statistical analysis report. NCES, National Center for Education Statistics, U.S. Department of Education, Institute of Education Science, 2006-328.

Fredricks, J. A., Blumenfeld, P. C. & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research.* 74, 59-109.

Garris, R., Ahlers, R. & Driskell, J. E. (2002). Games, motivation, and learning: a research and practice model. *Simulation & Gaming.* 33, 441-467.

Habgood, M. & Ainsworth, S. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences.* 20(2), 169-206.

Hughes, K. (2010). Educational software for off-task behavior. (Unpublished Major Qualifying Project). Worcester Polytechnic Institute, MA.

Karweit, N. & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of Educational Psychology,* 74(6), 844-851.

Kraemer, E., Davies, S., Arndt, K. & Hunley, S. (2012). A comparison of the mystery motivator and the *get 'em on task* interventions for off-task behaviors. *Psychology in the schools.* 49(2), 163-175.

McKevitt, B. C., Dempsey, J. N., Ternus, J. & Shriver, M. D. (2012). Dealing with behavior problems: The use of positive behavior support strategies in summer programs. *Afterschool Matters.* (15), 16-25.

Pardos, Z., Baker, R.S.J.d., San Pedro, M.O.Z., Gowda, S.M. & Gowda, S. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge.* 117-124.

Puustinen, M., Volckaert-Legrier, O., Coquin, D. & Bernicot, J. (2008). An analysis of students' spontaneous computer-mediated help-seeking: A step toward the design of ecologically valid supporting tools. *Computers & Education*. 53(4), 1040-1047.

Roll, I., Aleven, V., McLaren, B.M. & Koedinger, K.R. (2007). Can help seeking be tutored? Searching for the secret sauce of metacognitive tutoring. *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, 203-210.

Rowe. J., McQuiggan, S., Robison & J., Lester, J. (2009). Off-Task Behavior in Narrative-Centered Learning Environments. *Proceedings of the 14th International Conference on AI in Education.* 99-106.

Roscoe, R., Segedy, J., Sulcer, B., Jeong, J. & Biswas, G. (2013). Shallow strategy development in a teachable agent environment designed to support self-regulated learning. *Computers & Education*. 62, 286-297.

Sabourin, J., Rowe, J., Mott, B. & Lester, J. (2011). When Off-Task in On-Task: The Affective Role of Off-Task Behavior in Narrative-Centered Learning Environments. *Proceedings of the 15th International Conference on Artificial Intelligence in Education*. 534-536.

Sabourin, J., Rowe, J., Mott, B. & Lester, J. (2013). Considering Alternate Futures to Classify Off-Task Behavior as Emotion Self-Regulation: A Supervised Learning Approach. *Journal of Educational Data Mining*. 5(1), 9-38, 2013.

San Pedro, M.O.Z., Baker, R.S.J.d., Bowers & A.J., Heffernan, N.T. (2013). Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. *Proceedings of the 6th International Conference on Educational Data Mining*. 177-184.

San Pedro, M.O.C., Baker, R., Rodrigo, M.M. (2011) Detecting Carelessness through Contextual Estimation of Slip Probabilities among Students Using an Intelligent Tutor for Mathematics. *Proceedings of 15th International Conference on Artificial Intelligence in Education*. 304-311.

Schwartz, D., Chase, C., Chin, D., Kwong, H., Oppezzo, M., Okita, S., Roscoe, R., Jeong, H., Wagster, J. & Biswas, G. (2009). Interactive metacognition: Monitoring and regulating a teachable agent, in D.J. Hacker, J. Dunlosky & A.C. Graesser (Eds.), *Handbook of Metacognition in Education. New York, NY: Routledge.*

Sayeski, K. L. & Brown, M. R. (2011). Developing a classroom management plan using a tiered approach. *Teaching Exceptional Children. 44*(1), 8-17.

Sottilare , R. (2013) Personal communication.

Verginis, I., Gouli, E., Gogoulou, A. & Grigoriadou, M. (2011). Guiding learners into reengagement through the SCALE environment: An empirical study. *IEEE Transactions on Learning Technologies. 4*(3), 275-290.

Villalta, M., Gajardo, I., Nussbaum, M., Andreu, J. J., Echeverría, A. & Plass, J. L. (2011). Design guidelines for classroom multiplayer presential games. *Computers & Education*.   57(3), 2039-2053.

Wang, M. S. & Eccles, J. (2012). Adolescent Behavioral, Emotional, and Cognitive Engagement Trajectories in School and Their Differential Relations to Educational Success. *Journal Of Research On Adolescence. 22*(1), 31-39.

Weiss, M. J., DelPizzo-Cheng, E., LaRue, R. H. & Sloman, K. (2009). ABA and PBS: The dangers in creating artificial dichotomies in behavioral intervention. *Behavior Analyst Today. 10*(3), 12.

Wixon, M., Baker, R.S.J.d., Gobert, J., Ocumpaugh, J. & Bachmann, M. (2012) WTF? Detecting Students who are Conducting Inquiry Without Thinking Fastidiously. *Proceedings of the 20th International Conference on User Modeling, Adaptation and Personalization (UMAP 2012)*. 286-298.

Chapter 5 – The Importance of Narrative as an Affective Instructional Strategy

Mark O. Riedl[1] and R. Michael Young[2]
[1] School of Interactive Computing, Georgia Institute of Technology;
[2] Department of Computer Science, North Carolina State University

Introduction

Virtual simulations, computer games, and other game-like virtual worlds are progressively being adopted for "serious" purposes, such as training and education. Computer-based training systems and games share many similarities: both involve a progression of skill-based activities of increasing complexity and difficulty. The learner is expected, over the duration of the game, to master a set of skills. Skill mastery is one of the fundamental principles behind the success of computer games; indeed, this mastery of the game is a fundamental aspect of having fun in computer games (Koster, 2005). It is easy to see the appeal of games and game-like virtual environments for the purposes of training and education. The skill-based progression typically used in computer games and the desire for players to achieve mastery of particular game-relevant skills can be mapped to educational outcomes and pedagogy.

In computer games, it is not always enough to have a progression of skill-based activities. Many game genres use fictional context to reinforce the immersion within the game world and motivate the skill-based activities. These fictional contexts answer the question "why am I, as the player, engaging in a particular activity?" The fictional context may further induce an affective response from the player: dramatic tension over how events are unfolding, strong positive or negative feelings toward virtual characters, or suspense over what might happen next. In many games, skill-based activities are often structured through narrative, mission, quest, or scenario. These narratives, missions, quests, and scenarios can manifest as backstory or non-skill-based interactive game play such as moving through a virtual environment and interacting with virtual, non-player characters. Without narrative context, educational and training games may be perceived as a progression of drills without a purpose other than mastery itself. The use of fictional context is one possible way to more fully engage learners and motivate them to partake in skill-based progressions. In this chapter, we use the term *narrative* to mean a predetermined, temporally ordered set of actions or events. To that end, missions, quests, and scenarios are forms of narrative because they involve a temporally ordered sequence of events.

Unfortunately, the similarities between computer games and game-like learning environments are sometimes only skin-deep when it comes to the use of narrative contexts to motivate and engage. In particular, it is often sufficient for entertainment-based computer games to rely on linear narrative sequences and linear skill-progressions. Typically, an important design consideration in entertainment-based games that all players have the same experience. However, intelligent educational and training technologies require the ability to adapt to deliver the right educational content to the right learner at the right time. That is, variability in skill ability and rate of mastery must be accounted for because a skill-progression that is too slow or too fast may result in the learner abandonment. Research in ITSs has made significant gains in understanding how to model learner abilities and deliver the right problem to individual learners in the right sequence. ITSs can be broadly described as implementing two nested processes (VanLehn, 2006). The inner process is one of recognizing learner difficulties when solving a problem and selecting the most effective means of feedback and remediation. The outer process is one of selecting the best problem for the learner to work on next. The next problem the learner works on may come from a library of problems or may be automatically generated based on the learner's immediate needs and abilities. If an educational or training game must dynamically select skill-based events in order to maxim-

ize player learning potential, then a single, linear narrative may no longer be sufficient and artificial intelligence—in the form of automated story generation—can be brought to bear to construct new narratives that motivate and contextualize the learner-customized skill progression.

In this chapter, we describe two systems that use computational models of narrative generation to create game play experiences that directly support the learning process. One system, Annie, uses a model of a game's task domain to track players' knowledge and alter a player's challenges as they demonstrate mastery of or misconceptions around a particular skill. Another system, Game Tailor, automatically generates a sequence of skill challenges through which a player will progress and then creates a customized storyline in which the challenges are naturally embedded. These two systems demonstrate the effective role that explicit narrative models can play in the generation of tailored learning experiences within games.

Background and Related Research

A *narrative* is a predetermined, temporally ordered set of actions or events. Actions can be executed immediately in the virtual environment, whereas an event is a discrete period of time in which the actions of a number of characters are thematically or semantically related. An example of an event in a computer game is "player fights boss opponent" in a role-playing game or "user deletes a malicious virus program from memory" in a game-based virtual simulation of a computer. Each event may consist of a number of actions such as combat attacks, discourse acts, manipulation of file permissions, etc. The simplest narrative is a single, linear sequence. However, temporal ordering can support more sophisticated narrative structures, such as parallel actions and events. In this section, we provide an overview of the ways that story is treated in computer games, discuss artificial intelligence techniques for generating and adapting narrative structures, and compare narrative adaptation with ITSs.

Narrative in Computer Games

In computer games, narrative is used to motivate player behavior and establish the context for why the player is to perform certain activities. The narrative acts as an explanation, or context, for the activities the player is about to perform. The events making up a mission, quest, or scenario may be categorized as *skill-based* or *non-skill-based*. Skill-based events are periods of game play that require the player to attempt to perform a skill that is valued by the game designer or instructor. For entertainment-based computer games, skills may include finding and collecting items, solving puzzles, navigating through mazes, combat with opponents, etc. Non-skill-based events are periods of game play that do not require a skill valued by the game designer or instructor. Additionally, the game may involve periods of non-interactivity where the player is watching a cut-scene or in which the player's avatar is temporarily controlled by a script. Game designers use non-skill-based periods of play and non-interactive periods to advance a story, to create context for the next period of skill-based play, and motivate the player to achieve certain goals.

One aspiration of game design is to encourage players to move along an intentionally circuitous route to incorporate experiences that positively affect the player's enjoyment. Game designers refer to the circuitous route as the *golden path* and the most direct route as the *spine* (Bateman, 2007). Most games have a linear spine, providing little or no variation in which events are involved in completing the game. In contrast, the golden path contains additional, non-mandatory game elements that enhance other aspects of the player's experience. Perhaps equally important, however, the golden path enhances the player's sense of agency over the events that occur during game play and helps disguise the essentially static structure of the underlying spine.

One can visualize the structure of many computer games as a "string of pearls" where the pearls are periods of interactive game play (often referred to as levels or maps) and the string that holds all of the pearls together and space the pearls out is non-interactive narrative. From an implementation perspective, pearls are often implemented as sandbox environments – bounded simulations where possible actions are dictated by the underlying rules and physics of the game. Pearls may be skill-based or non-skill-based. The player engaged with a pearl until he or she triggers the conditions necessary to exit the pearl. This may involve reaching the end of a level, performing a particular action, or successfully demonstrating one or more skills. The narrative string that follows the most recently played pearl then sets up the context for the next successive pearl. Typically, successive pearls will require the player to demonstrate skills under more challenging circumstances. We refer to this as *skills progression*.

Story Generation, Interactive Narrative, and Intelligent Tutoring

Computer games for education and training differ from entertainment-based games in that game play by learners may be mandatory if the game is part of an educational curriculum or part of a training regimen, or required for skill or knowledge assessment. When this is the case, a string of pearls design will be insufficient because learners with different abilities or rates of mastery may require that learners experience different skill progressions. Variability in learner ability and rate of mastery may mean a fixed skills progression may be too difficult or too easy to promote effective learning (Vygotsky, 1978).

Artificial intelligence can be used to model learner skills and mastery rate, and produce tailored skills progressions that directly address learners' needs and abilities. This idea is not new; ITS researchers have sought to tailor learning environments to individuals. VanLehn characterizes an ITS as a process involving two nested loops. The outer loop performs problem generation, creating or selecting the next problem based on information about the learner, including traits, learning goals, and needs. The inner loop closely monitors every action the learner takes while performing the given task and uses this information to update a model of the learner and provide directed feedback. Zook et al. (2012) note that if an intelligent system can produce new narrative structures, then it may serve the purpose of problem generation while simultaneously contextualizing the learner's behaviors through narrative. Any performance-based feedback operating during skill-based events of the narrative can be thought of as equivalent to inner-loop remediation.

Automated story generation is the problem of automatically selecting a temporally ordered set of events that meet a set of criteria and can be told as a story. For story generation, there are two problems that one must address. The first is to computationally model narrative structure. The consensus among psychologists and computer scientists alike is that a narrative can be modeled as a semantic network of concepts (Trabasso, Secco & van den Broek, 1984; Graesser, Lang & Roberts, 1991; Young, 1999). Nearly all cognitively inspired representations of narrative rely on causal connections between story events. The second problem is to computationally model the narrative creation process and develop algorithms that implement the model. Approaches to automated story generation include simulation, planning, case-based reasoning, and natural language processing (NLP). The simulation approach (Meehan, 1976; Aylett et al., 2005; Cavazza, Charles & Mead, 2002) situates autonomous virtual agents in an environment and records their actions. One of the critiques of simulation approach is that coherent narrative sequences may not necessarily always emerge. To solve issues of narrative coherence, planning – the search for a sound and complete sequence of actions that achieves a goal situation – techniques have been developed that observe global structural patterns (Lebowitz, 1987, Porteous & Cavazza, 2009; Riedl & Young, 2010), employing cognitive models (Riedl & Young, 2010), or specialized heuristics and constraints (Porteous & Cavazza, 2009; Riedl, 2009). Case-based reasoning approaches to story generation reuse existing stories in new contexts (Turner 1994; Pérez y Pérez & Sharples, 2001; Gervás et al., 2005; Riedl, 2010). The

NLP approach to story generation is to mine word tuples or sentences from blogs (Swanson & Gordon, 2008) or text corpora (McIntyre & Lapata, 2009).

Interactive narrative (also *interactive storytelling* or *interactive drama*) is a form of digital entertainment in which users create or influence a dramatic storyline through actions, either by assuming the role of a character in a fictional virtual world or by issuing commands to autonomous, virtual non-player characters. The simplest interactive narratives, such as Choose-Your-Own-Adventure books and hypermedia, do not require artificial intelligence. A branching story graph is a directed graph where nodes are events and arcs are annotated with actions that the player can choose that lead to different narrative continuations. Branching story graphs can be manually authored, or procedurally generated by a story generator. Riedl and Bulitko (2013) provide an overview of AI approaches to interactive narrative. Mott et al. (1999) observe that narrative can be a useful tool for framing educational problem-solving activities. Interactive narrative has been explored as means of guiding humans through educational experiences and training scenarios (c.f., Rowe et al., 2011; Riedl et al., 2008; Magerko, Stensrud & Holt, 2006; Johnson & Valente, 2009; Marsella, Johnson & LaBore, 2000; Aylett et al., 2005; Thomas & Young, 2010).

Discussion

In the following sections, we describe two ways in which interactive narrative and story generation can support learners through remediation of misconceptions, generation of skills progressions, and contextualization of activity in the virtual world. First, we describe a system called Annie; Annie detects and addresses misconceptions about procedural knowledge. Because many computer games provide sandbox-style, exploratory environments for learning, players and learners typically have wide latitude to select actions and compose plans to achieve their in-game objectives. By design, these games provide many possible ways for a learner to navigate the task space. This presents an intelligent tutor with a challenge that Annie is intended to address: a game-based ITS must track the learner's plan and take action on-the-fly to remediate any misconceptions. Second, we describe Game Tailor, a system that addresses problem generation for serious games. Game Tailor determines the next skill in a skill progression that a learner should practice. Unlike tutoring systems that select the next new problem from a library, Game Tailor generates an entire skills progression at once and then generates a storyline that motivates all the problems the learner will work on in the sandboxes.

Annie: Leveraging Plan-Based Models of Narrative to Detect and Address Misconceptions

Exploratory environments provide students with freedom to choose different courses of action. This complicates the tutor's ability to know what the student it trying to do, which introduces uncertainty in knowing whether or not a student has a misconception about the domain. When the tutor decides a misconception exists, it is difficult to know when is the right time to provide support to remediate that misconception, as the student may have changed focus to a different task. As Van Joolingen, De Jong, and Dimitrakopoulou (2007) note, it is difficult to balance guidance with student exploration.

In our previous work on the Annie system (Thomas & Young, 2010; Thomas & Young, 2011), we have addressed these problems by leveraging a well-understood computational model of actions and the causal relationships between them used in automated planning. The style of action descriptions invented for the STRIPS system (Fikes & Nilsson, 1971) has continued to form the basis of much subsequent research in automated planning. Building on several distinct approaches to integrating automated planning with game domains (Mott & Lester, 2006; Mateas & Stern 2005; Cavazza, Charles & Mead, 2002; Riedl, Saretto & Young, 2003), the Annie system leverages a general plan-based knowledge representation intended both

to characterize a game-based learning environment's task domain as well as the knowledge of the tasks held by a learner.

STRIPS-style plan representations characterize actions available in a task domain schematically, defining an action in terms of its act-type, a set of *preconditions*, and a set of *effects*. Preconditions are logical terms that indicate just those conditions in the task domain that must be true in order for the action to execute correctly, while effects indicate all the ways that a task domain changes as a result of the successful execution of an action. As an example, consider a task domain within a game world focused on teaching users how remove malware from a PC. One action in this domain might be named deleteFile, corresponding to the action of deleting a file from the PC's hard drive. This action would have two parameters: one indicating the character or player initiator of the task and one naming the file to be deleted. Its preconditions would indicate that, before this action can be carried out, the file must exist and must not be in use. Further, the character performing the action must be limited to the player (e.g., no non-player character in the game can delete files). The effects for deleteFile would indicate that once the action succeeds, the file will no longer exist.

To build the model for what the student knows about deleting files, Annie begins by automatically deriving a set of meta-conditions from the known features of the deleteFile operator. The simplest model of the student's knowledge of the operators in the domain would register whether the student *knows* that a term appears as a precondition or an effect of a given action. For instance, Annie can generate requirements that a student knows that a file being deleted must exist, that it cannot be in use at the time, and that once the deleteFile action is performed, the file will no longer exist.

This simple approach to model construction fails to capture the uncertain nature of student knowledge in an exploratory environment where the student's understanding of the world evolves gradually. To represent this uncertainty, we employ a rough-grained five-valued scale (HighlyLikely, Likely, Neutral, Unlikely, HighlyUnlikely) to represent varying estimates of the likelihood that the student believes or knows about a particular facet of the domain, where "Neutral" is the default initial value.

To illustrate, in a game that teaches the processes involved in aerobic cellular respiration, Annie may observe a student behavior that implies that the student knows an effect of the Krebs cycle is the production of CO_2 waste but may have no information yet on whether the student knows another effect of the process is the production of H_2O. This could be represented in the student model by marking the hasEffect condition corresponding to CO_2 production of a particular action in the Krebs cycle as HighlyLikely, while the effect that produces H_2O is marked as a student belief with Neutral likelihood.

Like many ITSs, Annie's core tutorial reasoning is situated in a loop interleaving student and system-controlled actions. Each time an action is taken in the world, either by the student or the system, Annie updates its student model by consulting a library of general diagnostic templates. These templates encode domain-independent plan reasoning diagnostics such as cases where a student seems to be ignorant of a precondition of a particular action. For example, if a student attempts an action for which some of the preconditions are not satisfied, a rule in one of these diagnostic templates fires to update the student model by lowering its confidence that the student is aware of those preconditions.

Annie uses the updated student model in consulting a second domain-independent library containing remediation templates that can be used to generate scaffolding. For example, if the plan shows that a particular task must be performed for the student to make progress toward plan goals, and Annie notes particular gaps in the student model pertaining to that action (e.g., student has an incorrect model of its effects), it will prompt the student about that action.

As mentioned above, execution loops are common in ITSs that often operate with nested loops, one iterating over problems or tasks and another, nested within it, operating over individual steps in the problem. As described here, Annie's loop is often focused on individual steps in a task. Unlike a concentric loop architecture, however, Annie is free to switch to a completely different higher-level task or problem as a student interleaves tasks within the game.

A potentially difficult paradox for Annie's design is that as the student progresses, Annie gains more and more information about the state of the student's knowledge, but has less and less time remaining to act on these inferences. In order to characterize how close a student is to achieving important goals or milestones within a game world, we leverage the planning-based representation of the game world's task domain to compute the game world's *plan space* – a directed graph that characterizes the space of all possible plans for achieving a given set of goals in a specific game world. Planning algorithms called plan-space planners (Kambahmpati, Knoblock & Yang, 1995) construct plan spaces as part of their search process when solving a planning problem. Because the proper sequencing of actions within a plan relies on valid student knowledge regarding the tasks involved, Annie can use the plan space it constructs to prioritize and sequence its strategies for guiding the student toward acquiring the requisite knowledge.

A plan-based representation can provide a language simultaneously describing learning content and game play. With automated planning techniques, we can ensure that the spine of the game is traversed, while encouraging the player to explore far beyond the small set of detours built into a golden path. Through planning, a widely varied golden landscape unfolds where individual users can explore a variety of experiences tailored to their particular educational and entertainment aspirations.

Recapitulating Game-Based Learning Through Planning

Gee (2003) described a rich set of learning principles evident in commercial games and Quintana et al. (2004) described a framework that identified many of the scaffolding techniques used in exploratory ITS research, but neither of these descriptions lends itself to a generative model. Each leaves it to the artistic spirit of game or tutorial designers to decide when, where, and how extensive the computational support should be. Annie, however, requires a generative model for game-based learner guidance. We have built such a model inspired by the descriptions of Gee and Quintana, providing the following capabilities:

1) Each learning principle is articulated through one or more plan-based templates to allow automatic generation of game play elements that embody that principle.

2) Generation is performed at run-time, allowing the game to dynamically adapt to the behaviors exhibited by the student.

3) Systems can measure or specify the frequency and extent to which learning principles are realized. In other words, the model provides researchers with a mechanism to freely vary the prevalence of one principle *vs*. another and measure the effects.

Nine of the 36 learning principles articulated by Gee were selected as initial candidates for testing this generative model. Three of these are described briefly here.

Overt telling is kept to a well-thought-out minimum, allowing ample opportunities for the learner to experiment and make discoveries.

We use the term *remediation* to describe an action Annie inserts into the game environment to attempt to correct what it perceives to be a misapprehension on the part of the student. We can count the number of remediations applied for each student, the best case, worst case and average number of remediations

required for each particular knowledge component, and the comparative frequency of stronger or weaker hints that correspond to different type of remediations. Across a broad range of students, these measurements can be used to characterize the difficulty of different parts of the game world and help pinpoint areas where more student guidance opportunities may be required.

Remediations are organized in such a way as to allow Annie to choose between successively more explicit modes of instruction. This builds on extensive ITS research into the optimal selection strategy between the frequently used guidance options of 'Prompt', 'Hint', 'Teach', or 'Do'.

There are multiple ways to make progress or move ahead. This allows learners to make choices, rely on their own strengths and styles of learning and problem-solving, while also exploring alternative styles.

Annie can quantify the number of distinct successful plans, the number of qualitatively different plans in the plan space, the number of actions that must be included in any successful plan, or even the ratio of the number of these critical actions to the mean total number of actions in successful plans.

Annie allows for extensive mining of the space of potential plans to reveal bottlenecks, potential for off-task activity, etc., in a way that could be much cheaper and more extensive than traditional game design play testing strategies.

The learner is given explicit information both on-demand and just-in-time, when the learner needs it or just at the point where the information can best be understood and used in practice.

The timeliness of explicit information can be measured by the duration of the interval between when the information is provided and when it is needed. This can be compared and contrasted with the number of opportunities for on-demand information in the environment. For some students or groups of students, Annie may want to vary how far in advance help can be provided based on projected memory persistence of those students. As post-hoc measurements, analysis of these properties over many students can be used to calibrate guidance within Annie.

Advantages of Plan-Based Game Design

Our intention with the development of the Annie system was to demonstrate that a nominal plan-based knowledge representation can lead to a computational framework that can automatically synthesize and adapt gameplay/teaching at an atomic level. In this work, we selected a set of learning principles and leveraged a plan-based design to realize these principles in arbitrary domains. Specifically, our knowledge representation synthetically generates game structures that implement these principles, requiring less time, and money, resulting in a shorter and cheaper development cycle. Because these structures are automatically generated, their instantiation can be shifted to run-time, so they can be tailored to the immediate and subtle learning needs of the individual rather than the statically defined and obvious extremes of an entire population. Finally, the rules governing how and when to change course are visible and modifiable, rather than entwined with tutorial algorithms. This enables the system to conform to externally specified metrics for particular applications.

The use of a plan-based knowledge representation breaks the game spine into interchangeable parts, allowing for dynamic synthesis of game progression while ensuring that the player eventually traverses segments of the spine nominated as particularly critical. Any fixed branching structure could be implemented through a plan-based representation by representing each critical action choice as a distinct operator with unique prerequisites and effects. But planning not only replicates the expressivity of existing game progression, it allows for a much wider variety of scaffolding techniques, partial-ordering

of actions, and varied bindings of particular game elements and arbitrary number of repetitions or cycling through particular types of actions.

Game Tailor: Generating and Contextualizing Skills Progressions

Problem generation assesses the question of what problem the learner should work on next. Serious games can take a lesson from entertainment-based games by using an unfolding plotline to motivate problems and create affective engagement with content. In computer games, the skills progression is an important part of creating a sense of mastery and fun. Game Tailor creates a skills progression as a sequence of skill-based events (sandboxes) that is tailored to an individual player and provides a storyline that sets up and explains the skill-based events.

Challenge tailoring is the problem of matching the difficulty of skill-based events over the course of a game to a specific player's abilities. While not strictly narrative generation, we first consider the problem of generating a skills progression tailored to individual player abilities. This is analogous to the creation of a string of skill-based pearls, but without the narrative "string" that ties the skill-based events together. Once we know the sequence of skill-based events that a player will encounter, the next step is to generate the narrative string that contextualizes each skill-based event. We emphasize the selection of the right sequence of skill-based events for the right player at the right time. Although our approach to challenge tailoring is applicable to a number of serious games, we will illustrate our approach through a simple combat game inspired by The Legend of Zelda. In The Legend of Zelda, the player must lead a team of avatars into periodic combat with teams of opponent monsters. In such a game challenge tailoring may manifest as configuring the number, health, or damage dealt by various enemies at various times through-out the game. CT is similar to Dynamic Difficulty Adjustment (DDA), which only applies to online, real-time changes to game mechanics to balance difficulty. In contrast, CT generalizes DDA to both online and offline optimization of game content and is not limited to adapting game difficulty. Challenge contextualization is the problem of constructing a chain of non-skill-based events and/or non-interactive sequences that set up the conditions for skill-based events and motivate their occurrence to the player. For example, the challenge of slaying a dragon may be contextualized by the dragon kidnapping a princess.

Challenge Tailoring

Realizing challenge tailoring requires both a player model and an algorithm to adapt content based on that model. Effective player modeling for the purposes of challenge tailoring requires a data-driven approach that is able to predict player behavior in situations that may have never been observed. Because players are expected to master skills over time when playing a game, the player model must also account for temporal changes in player behavior, rather than assume the player remains fixed. Modeling the temporal dynamics of a player enables an adaptive game to more effectively forecast future player behavior, accommodate those changes, and better direct players toward content they are expected to enjoy. Further, forecasting enables player models to account for interrelations among sequences of experiences—accounting for how foreshadowing may set up a better future revelation or how encountering one set of challenges builds player abilities to overcome related challenges that build off of those. We employ tensor factorization techniques to create temporal models of objective player game performance over time. We demonstrate the efficacy of the approach below in a turn-based role-playing game. Further details and evaluation can be found in Zook and Riedl (2012).

Tensor factorization techniques decompose multidimensional measurements into latent components that capture underlying features of the high-dimensional data. Tensors generalize matrices, moving from the two-dimensional structure of a matrix to a three or more dimensional structure. For our player modeling approach, we extend two-dimensional matrices representing player performance against particular enemy

types to add a third dimension representing the time of that performance measure. Tensor factorization is an extension of matrix factorization, which offers the key advantage of leveraging information from a group of users that has experienced a set of content to make predictions for what a new group of individuals that has only been partially exposed to that content will do. Specifically, if matrix factorization represents user data as a matrix $M = U \times I$ indicating user preference ratings on items, then tensor factorization represents user data as a matrix $Z = U \times I \times T$. Both approaches extract latent factors relating to users and items (and time). The latent factors extracted from the matrix are used to predict missing user ratings of items. The technique for extracting latent factors from the matrix is beyond the scope of this chapter (*c.f.*, Zook & Riedl, 2012). In our usage of tensor factorization, items are challenges—combat, puzzles, or problems to be solved—and ratings are measures of player performance. While we believe our work is the first application of tensor factorization to challenge tailoring problems, we note that similar techniques have been used to model student performance over time on standardized tests (Thai-Nghe, Horvath & Schmidt-Thieme, 2011).

In our turn-based combat domain, the player leads a team of hero characters against a team of opposing monsters. Each combat is a single skill-based event in a skills progression involving a number of combats. The player can cast a number of spells and different spell types work against different types of monsters. While the role-playing game is a good demonstration of challenge tailoring, it is also a skill learning task. We intentionally created a spell system that was difficult to completely memorize, but contained intuitive combinations—water spells are super-effective against fire enemies—and unintuitive combinations—undeath spells are super-effective against force enemies—ensuring that skill mastery could only be achieved by playing the game. Players do not do well if they do not learn from experience, the effectiveness of spells against different opponents. More complicated domains in which the learner must correctly perform complex procedures—such as those used by Annie—are also possible.

We model performance instead of difficulty because performance is objectively measurable while difficulty is subjective. Difficulty and performance have been shown to be significantly (inversely) correlated in the domain of turn-based combat (Zook & Riedl, 2012). Tensor factorization tends to outperform matrix factorization by taking into account the rate at which the player learns the skill of effectively casting spells against opponents of different types. That is, it can predict the actual effectiveness of a player many combats into the future after training. Accuracy of the model is dependent on (a) the number of combats observed of a given individual and (b) the number of overall users represented in the tensor. We find that for our simple combat game, we can achieve high accuracy with as few as 6 training examples per individual and as few as 30 different players. However, more complicated games will require a larger database of player traces. Fortunately, matrix and tensor factorization spreads the model training over a large number of users such that the system need only observe a small number of ratings per user.

To generate particular skill progression of combat episodes, the system uses an author-defined performance curve. Typically, a performance curve presents the player with a smooth increase in difficulty, i.e., a decrease in player performance over time. Other curves are possible. For example, a curve expressed by $p = c$ (a horizontal line at a fixed constant, c) indicates a game in which the difficulty appears to remain the same, even as the player's skills improve. A dramatic arc, in which the player progressively faces more and more dire challenges until the toughest challenge is overcome and difficulty eases off, can be created with a U-shaped curve. More complicated patterns, such as a series of rises and falls, can express complex designer intentions.

Skills progression generation is an optimization process in which skill-based events are selected such that distance between the predicted performance of the individual on the skills of each event and the performance curve is minimized. A variety of techniques may be applied to solve this dynamic optimization problem including constraint satisfaction, dynamic programming, and heuristic search techniques such as

genetic algorithms (Smith and Mateas, 2011; Togelius et al., 2011; Sorenson, Pasquier, and DiPaola, 2011). In contrast to the reactive, near-term changes typically employed in DDA (Magerko, Stensrud, and Holt, 2006; Hunicke and Chapman, 2004), temporal player models are able to also proactively restructure long-term content to optimize a global player experience. Our technique selects sets of enemies for each skill-based event automatically through combinatorial optimization using Answer Set Programming (Baral, 2003). Answer Set Programming is a declarative programming language used for finite domain constraint solving using logic programming semantics.

Challenge Contextualization

But why is the player engaging in the activities that require skills to be practiced? While the sequence of skill-based events can be considered a narrative, the transition from skill-based event to skill-based event creates the context necessary for the player to understand how the skill-based events fit together. Challenge contextualization addresses the issue of player motivation by embedding the skills progression into a larger narrative that does not directly challenge the learner, but engages the learner via fictional means. Challenge contextualization is a form of narrative generation. While challenge tailoring and challenge contextualization can be performed in parallel, we assume a tailored sequence of skill-based events already exists; the selection and parameterization of skill-based events takes precedence in serious games. Thus, the narrative generation problem becomes one of selecting and spacing all skill-based events before "filling the gaps" with non-skill-based, contextualizing events.

Planning is one of the most common approaches to story generation. Planning is the search for a sequence of operations – in this case, events – that transform the world from an initial state into one in which a goal situation holds. To apply story planning to challenge contextualization, the goal situation must be such that it is achieved only if the conditions necessary to establish each skill-based event in turn are achieved at some point in the plan and in order. Skill-based events are sandboxes, and while the actions that occur within a sandbox simulation is dependent on the player and therefore uncertain, all sandboxes have an initial condition (e.g., player and enemy are co-located in the virtual world; a computer has become infected with malware) and a terminal condition (e.g., the opponent is dead; the computer is free of malware).

There are two ways of using story planning in challenge contextualization. The first is to produce distinct planning problems for each pair of skill-based events. In the first iteration of this technique, the initial state is the initial state of the world as specified by a game designer, and the first goal situation is the initial conditions of the first skill-based event. In subsequent iterations, the initial state is the world state that results from executing the plan from the prior iteration updated with the terminal conditions of the last skill-based event, and the goal situation will be the initial conditions of the next skill-based event. The advantage of this approach is that the planning problems are smaller, and therefore, more tractable. The disadvantage is that narrative decisions made in prior iterations become locked-in and cannot be changed if it is later discovered that it is impossible or awkward to fill a later gap between two skill-based events.

The second story planning approach to challenge contextualization is to consider the entire sequence of skill-based events as part of a single, larger planning problem. We cannot hope that a planner will serendipitously establish the conditions necessary for each skill-based event. To generate a single narrative plan, we must determine how to incorporate all skill-based events simultaneously. In planning, an *island* (Hayes-Roth and Hayes-Roth 1979) is a set of states through which the solution plan must traverse. Any sequence of operators that does not traverse through at least one state in an island at any point is pruned. The initial condition of each skill-based event is an island and each island must be traversed in the order determined by challenge tailoring. Riedl (2009) describes a technique for incorporating islands into partial order planning. Islands are represented as events with preconditions and effects. The initial plan is seeded with the islands, which are temporally ordered according to the skills progression generated

during challenge tailoring. Thus, the preconditions of each island become sub-goals that must be achieved by the planner by inserting non-skill-based events. The solution to the challenge contextualization planning problem is a sequence of events that interleave skill-based and non-skill-based events.

Together challenge tailoring and challenge contextualization provide a solution to the "problem generation" portion of ITSs that focuses on motivating the learner and creating affective and engagement through narrative. The narrative – in particular the non-skill-based events – is not strictly necessary, but breaks up the skill-based events and provides a *reason* for why skills and knowledge must be brought to bear on a sequence of increasingly difficult problems.

Recommendations and Future Research

Narrative is one of the fundamental modes for understanding the worlds around us, whether those worlds are real or virtual. Psychological studies show that narrative is read approximately twice as fast as informational text but remembered twice as well (Graesser, Olde & Klettke, 2002) so clearly it holds a distinguished status in the cognitive system. Virtual environments like computer games have come to blur the distinction between fictional worlds and everyday life as millions of people extend their daily social, leisure and professional identities into these contexts. To a great extent, these interactive systems rely on the explicit role that narrative plays in the design of their users' interactions for their effectiveness.

As intelligent systems develop the capability to model narrative and players' interactions within a narrative space, we argue that the capability to reason about and manipulate story structure in response to learner needs is critical. One key element to this capability is centered on a shift from current games' design focus of linear storylines to more open-ended exploratory environments. Annie's modular model of a game's task environment allows the system to track players as they explore the narrative space of a game and dynamically adjust the story content to address misconceptions as they are identified. The creation of tailored narrative experiences that provide individual learners with the right learning experience at the right time is generally intractable within the context of modern game design practices. For serious games to have the optimal impact on learning and mastery, the narrative experience must address both the pedagogical needs of the learner and encourage affective engagement with content, context to understand why problems are being solved, and motivation to work on progressively harder problems over a long duration of time. Game Tailor seeks to mask a progression of open-ended problem spaces as an unfolding plotline similar to those found in modern computer games while directly addressing the need for tailored pedagogical and narrative content.

Despite recent progress in remediation in open-ended exploratory environments, skills progression generation, and story generation, there are a number of future steps that will make for more robust, scalable, and affectively engaging experiences. First and foremost, automated story generation is a hard problem. While we have shown a considerable gain in computational story generation capabilities, story generation systems such as that used by Game Tailor still do not reliably create narrative structures that fully engage players and learners affectively. That is, automated story generation systems do not understand how the structures they generate produce affective responses in human readers, players, and learners. Recent work suggests that it may be possible for automated story generation systems to computationally model human affective responses to suspense (Cheong, 2007; O'Neill, 2013) and intentionally produce dramatic conflict between virtual characters (Ware et al., forthcoming). Second, the linkages between tasks and story are not always clear, nor easy to computationally model. More sophisticated generative models of task progressions are necessary that incorporate procedure and skill level (c.f., Andersen, Gulwani & Popović, 2013). But even this is not enough, sandboxes are simulations that support open-ended exploration and being able to embed a procedure, task, or skill into a virtual exploratory environment is still not well understood. To the extent that procedures can be represented as narra-

tives – albeit at the level of action instead of event – Hartsook et al. (2011) present initial steps toward dynamically creating open-ended virtual worlds that simultaneously support specific narrative elements and open-ended exploration. As these technologies progress, ITSs that exist within the context of serious games and interactive narratives will present learners with more immersive, more engaging learning experiences. By offloading many of the creative and pedagogical decisions onto intelligent systems embedded within these games, we may be able to reach larger populations of learners in informal and non-traditional learning environments.

References

Andersen, E., Gulwani, S & Popović, Z. (2013). A trace-based framework for analyzing and synthesizing educational progressions. *Proceedings of the 2013 International Conference on Human Factors in Computing Systems.*

Aylett, R., Louchart, S., Dias, J. & Paiva, A. (2005). FearNot! an experiment in emergent narrative. *Proceedings of the 5th International Conference on Intelligent Virtual Agents.*

Baral, C. (2003). *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge University Press.

Bateman, C. (2007). *Game writing: Narrative skills for videogames*. Charles River Media.

Cavazza, M., Charles, F. & Mead, S. (2002). Planning characters' behaviour in interactive storytelling. *Journal of Visualization and Computer Animation*, 13, 121-131.

Cheong, Y. (2007). *A Computational Model of Narrative Generation for Suspense*. PhD dissertation, North Carolina State University.

Fikes, R. & Nilsson, N. (1971). STRIPS: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2, 189-208.

Gee, J.P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment*, 1(1), 20–20.

Gervás, P., Díaz-Agudo, B., Peinado, F. & Hervás, R. Story plot generation based on CBR. *Journal of Knowledge-Based Systems*, 18(4-5), 235-242.

Graesser, A., Lang, K.L. & Roberts, R.M. (1991). Question answering in the context of stories. *Journal of Experimental Psychology: General*, 120(3), 254-277.

Graesser, A.C., Olde, B., and Klettke, B. (2002). How does the mind construct and represent stories? In M.C. Green, J.J. Strange & T.C. Brock (Eds.), *Narrative Impact: Social and Cognitive Foundations* (231-263). Mahwah NJ: Lawrence Erlbaum Associates.

Hartsook, K., Zook, A., Das, S. & Riedl, M.O. (2011). Toward supporting stories with procedurally generated game worlds. Proceedings of the 2011 IEEE Conference on Computational Intelligence in Games.

Hayes-Roth, B. & Hayes-Roth, F. (1979). A cognitive model of planning. Cognitive Science, 3.

Hunicke, R. & Chapman, V. (2004). AI for dynamic difficulty adjustment in games. Proceedings of the AAAI Workshop on Challenges in Game Artificial Intelligence.

Johnson, W.L. & Valente, A. (2009). Tactical language and culture training systems: Using AI to teach foreign languages and cultures. *AI Magazine*, 30(2), 72-83.

Kambhampati, S., Knoblock, C.A. & Yang, Q. (1995). Planning as refinement search: a unified framework for evaluating design tradeoffs in partial-order planning. *Artificial Intelligence*, 76(1-2), 167-238.

Koster, R. (2005). *A Theory of Fun for Game Design*. Paraglyph Press.

Lebowitz, M. (1987). Planning stories. *Proceedings of the 9th Annual Conference of the Cognitive Science Society*.

Magerko, B., Stensrud, B. & Holt, L. (2006). Bringing the schoolhouse inside the box—a tool for engaging, individualized training. *Proceedings of the 25th Army Science Conference*.

Marsella, S., Johnson, W.L. & LaBore, K. (2000). Interactive pedagogical drama. *Proceedings of the 4th International Conference on Autonomous Agents*.

Mateas, M. & Stern, A. (2005). Structuring Content in the Façade Interactive Drama Architecture. *Proceedings of the 1st AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*.

McIntyre, N. & Lapata, M. (2009). Learning to tell tales: A data-driven approach to story generation. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing*.

Meehan, J. (1976). *The Metanovel: Writing Stories by Computers*. PhD dissertation, Yale University.

Mott, B.W., Callaway, C., Zettlemoyer, L., Lee, S.Y. & Lester, J. (1999). Towards Narrative-Centered Learning Environments. In M. Mateas and P. Sengers (eds.) *Narrative Intelligence: Papers from the AAAI Fall Symposium* (Technical Report FS-99-01). AAAI Press.

Mott, B.W. & Lester, J. (2006). U-director: a decision-theoretic narrative planning architecture for storytelling environments. Proceedings of the 2006 International Conference on Autonomous Agents and Multi Agent Systems.

O'Neill, B.C. (2013). *A Computational Model of Suspense for the Augmentation of Intelligent Story Generation*. Ph.D. Dissertation, Georgia Institute of Technology.

Pérez y Pérez, R. & Sharples, S. (2001). MEXICA: A computer model of a cognitive account of creative writing. *Journal of Experimental and Theoretical Artificial Intelligence*, 13, 119-139.

Porteous, J. & Cavazza, M. (2009). Controlling narrative generation with planning trajectories: the role of constraints. *Proceedings of the 2nd International Conference on Interactive Digital Storytelling*.

Quintana, C., Reiser, B.J., Davis, E.A., Krajcik, J., Fretz, E., Duncan, R.G., Kyza, E., Edelson, D. & Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, 13(3), 337–386.

Riedl, M.O. (2009). Incorporating authorial intent into generative narrative systems. In S. Louchart, M. Mehta, and D. Roberts (eds.) Intelligent Narrative Technologies II: Papers from the AAAI Spring Symposium (Technical Report SS-09-06). AAAI Press.

Riedl, M.O. (2010). Case-based story planning: Creativity through exploration, retrieval, and analogical transformation. *Minds and Machines*, 20(4).

Riedl, M.O., Saretto, C.J. & Young, R.M. (2003). Managing interaction between users and agents in a multi-agent storytelling environment. Proceedings of the 2003 International Conference on Autonomous Agents and Multi Agent Systems.

Riedl, M.O. & Bulitko, V. (2013). Interactive narrative: An intelligent systems approach. AI Magazine, 34(1), 67-77.

Riedl, M.O., Stern, A., Dini, D. & Alderman, J. (2008). Dynamic experience management in virtual worlds for entertainment, education, and training. International Transactions on System Science and Applications, 3(1), 23-42.

Riedl, M.O. and Young, R.M. (2010). Narrative planning: Balancing plot and character. Journal of Artificial Intelligence Research, 39, 217-268.

Rowe, J., Shores, L., Mott, B. & Lester, J. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education*, 21, 115-133.

Smith, A. & Mateas, M. (2011). Answer set programming for procedural content generation: A design space approach. IEEE Transactions on Computational Intelligence and AI in Games, 3(3), 187–200.

Sorenson, N., Pasquier, P. & DiPaola, S. (2011). A generic approach to challenge modeling for the procedural creation of video game levels. IEEE Transactions on Computational Intelligence and AI in Games, 3(3), 229–244.

Swanson, R. & Gordon, A. (2008). Say anything: A massively collaborative open domain story writing companion. Proceedings of the First International Conference on Interactive Digital Storytelling.

Thai-Nghe, N., Horvath, T. & Schmidt-Thieme, L. (2011). Factorization models for forecasting student performance. Proceedings of the 4th International Conference on Educational Data Mining.

Thomas, J. & Young, R.M. (2010). Annie: Automated generation of adaptive learner guidance for fun serious games. IEEE Transactions on Learning Technologies, 3(4), 329 – 343.

Thomas, J. & Young, R.M. (2011). Dynamic guidance for task-based exploratory learning. Proceedings of the Fifteenth International Conference on Artificial Intelligence and Education.

Togelius, J., Yannakakis, G., Stanley, K. & Browne, C. (2011). Search-based procedural content generation: A taxonomy and survey. IEEE Transactions on Computational Intelligence and AI in Games, 3(3), 172–186.

Trabasso, T., Secco, T. & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N.L. Stein, and T. Trabasso (eds.), Learning and Comprehension in Text. Lawrence Erlbaum Associates.

Turner. S.R. (1994). The Creative Process: A Computer Model of Storytelling. Lawrence Erlbaum Associates.

Van Joolingen, W.R., De Jong, T. & Dimitrakopoulou, A. (2007). Issues in computer supported inquiry learning in science. Journal of Computer Assisted Learning, 23(2), 111–119.

VanLehn, K. The Behavior of Tutoring Systems. International Journal of Artificial Intelligence in Education, 16.

Vernieri, T.M. (2006). A Web Services Approach to Generating and Using Plans in Configurable Execution Environments. Master's thesis, North Carolina State University.

Vygotsky, L. (1978). Mind and Society: The Development of Higher Psychological Processes. Harvard University Press.

Ware, S.G., R.M. Young, Harrison, B. & Roberts, D.L. (forthcoming) A computational model of narrative conflict at the fabula Level. IEEE Transactions on Computational Intelligence and Artificial Intelligence in Games.

Young, R.M. (1999). Notes on the use of plan structures in the creation of interactive plot. In M. Mateas and P. Sengers (eds.), Narrative Intelligence: Papers from the AAAI Fall Symposium (Technical Report FS-99-01), 164-167. AAAI Press.

Zook, A. & Riedl, M.O. (2012). A temporal data-driven player model for dynamic difficulty adjustment. Proceedings of the 8th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment.

Zook, A., Riedl, M.O., Holden, H.K., Sottilare, R.A. & Brawner, K.W. (2012). Automated scenario generation: toward tailored and optimized military training in virtual environments. *Proceedings of the 7th International Conference on the Foundations of Digital Games*.

CHAPTER 6 – Personalized Content in Intelligent Tutoring Systems

Steven Ritter[1], Anne M. Sinatra[2, 3], and Stephen E. Fancsali[1]
[1] Carnegie Learning; [2] U.S. Army Research Laboratory; [3] Oak Ridge Associated Universities

Introduction

While most efforts to improve student academic outcomes have focused on instruction, recent research shows great potential for advancing student outcomes through a focus on "non-cognitive factors" such as motivation, beliefs about learning, and metacognitive skills (Farrington et al., 2012). While ITSs have successfully improved instruction by adapting to evolving learner knowledge based on cognitive factors (Pane, Griffin, McCaffrey & Karam, 2014), research has shown potential to magnify their effectiveness through incorporation of adaptive instruction based on better understanding of non-cognitive factors. For example, researchers have developed data-driven strategies to "detect" affective states like boredom and confusion from ITS log data (e.g., Baker et al. 2012). Other approaches deploy behavioral indicators like mouse movement (e.g., Sottilare & Proctor, 2012) and biometric or physical sensors like electromyography (EMG) (e.g., Conati, Chabbal & Maclaren, 2003; Conati & Maclaren, 2004) to infer learner mood and/or emotion in ITS environments. Another project led by Carnegie Learning seeks to develop a "hyper-personalized" intelligent tutoring architecture that will allow for tailoring the learner experience based on a wide variety of non-cognitive factors, including areas of learner interest outside of the classroom (Fancsali, Ritter, Stamper & Nixon, 2013). Still other research focuses on the integration of references to the self into learning materials (e.g., Sinatra, 2013). In this chapter, we focus on the latter two of these examples as non-cognitive factors upon which learning content (e.g., the text of mathematics word problems) can be personalized in ITSs. We begin by briefly describing a widely used ITS, Carnegie Learning's Cognitive Tutor® (CT) (Ritter, Anderson, Koedinger & Corbett, 2007) for mathematics and how it adapts to learners based on cognitive factors. We then introduce means by which a recent ITS based on the CT provides for personalization based on learner interest areas as well as recent work on solving logic puzzles including personalized content.

ITSs like the CT present students with complex multi-step problems. Cognitive task analysis and empirical methods such as Learning Factors Analysis (Koedinger, McLaughlin & Stamper, 2012) are used to identify particular knowledge components (KCs) or skills that are required to complete each step of the problem. The system is then capable of tracking student ability on each underlying KC step-by-step as the student solves each problem. Mathematics curricula in the CT are divided into topical units that are then divided into sections that treat particular sub-topics. Each section has a set of KCs with which it is associated and that the CT tracks for each student. When the CT judges that a learner has mastered all of the KCs associated with a section, the learner graduates to the following section (or unit, having graduated from all sections in a unit). Until a student graduates from a section, the CT adaptively selects problems for each learner, depending on the particular KCs that have yet to be mastered. Within a problem, hints and error feedback are adapted to particular students' problem-solving strategies and progress. Thus, all adaptation within this variant of CT depends on student knowledge and problem-solving states: cognitive factors.

Released in 2011, Carnegie Learning's middle school mathematics product, MATHia®, is based on the CT and, in addition to adaptively presenting problems based on cognitive factors, probabilistically "honors" learner interests in areas outside of the classroom (e.g., "sports & fitness," "arts & music") by presenting word problems that are tailored to such domains. MATHia also provides a facility whereby students can provide names of their classmates for inclusion in mathematics word problems (cf. Figure 1).

In this way, MATHia is able to personalize instruction based on non-cognitive factors (domain preferences and names of friends), in addition to the cognitive strategy usually employed in CT.
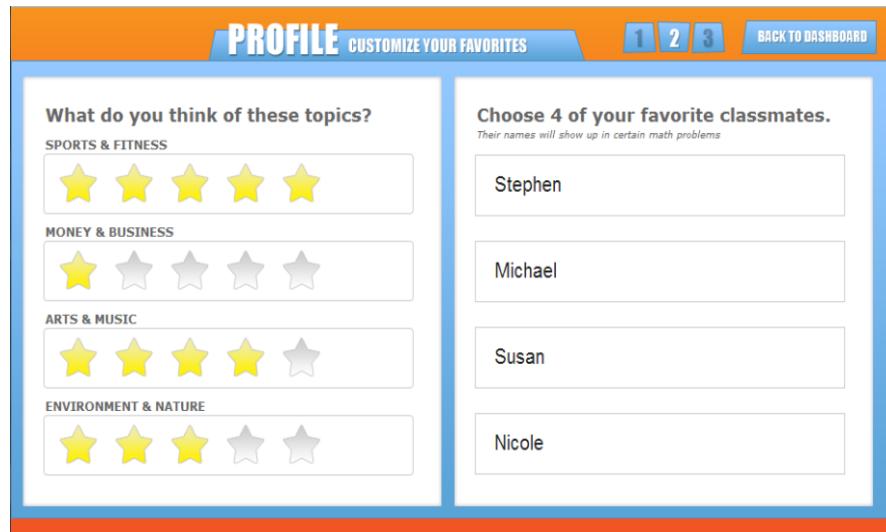
Figure 1. MATHia preferences profile.

The majority of work on personalization and tutoring has been in well-defined domains such as math and science. There are practical reasons for the large amount of research in these areas, as there are direct implications for student learning, and content in these areas can be more easily scored and adjusted than in less well-defined domains. One of the goals of GIFT is to allow for tutoring in both well-defined and ill-defined domains, and provide a set of authoring tools to assist instructors in developing the necessary assessments (Sottilare & Holden, 2013). Teaching skills such as deductive reasoning can have a long lasting effect on an individual's learning, but assessment of deductive reasoning skills is not as straightforward as performance on a math problem. Sinatra (2013) has recently explored the effects of personalization based on references to the self in the context of logic puzzle solving tutors implemented in the GIFT architecture.

Logic grid puzzles are complex puzzles that require individuals to use and apply deductive reasoning. They include a vignette, which sets up a story or context for the problem, as well as individual clues that assist the individual in narrowing down information and solving the puzzle. The inclusion of clues and story provides an opportunity for the puzzles to be personalized to an individual's interests or to include names with which the individual is familiar. GIFT 4.0, which was released in November 2013, includes a logic grid puzzle content domain, which contains a tutorial and assessments (e.g., multiple choice questions, clue questions, and an assessment puzzle). See Figure 2 for an example of a logic grid puzzle that is included with GIFT 4.0.
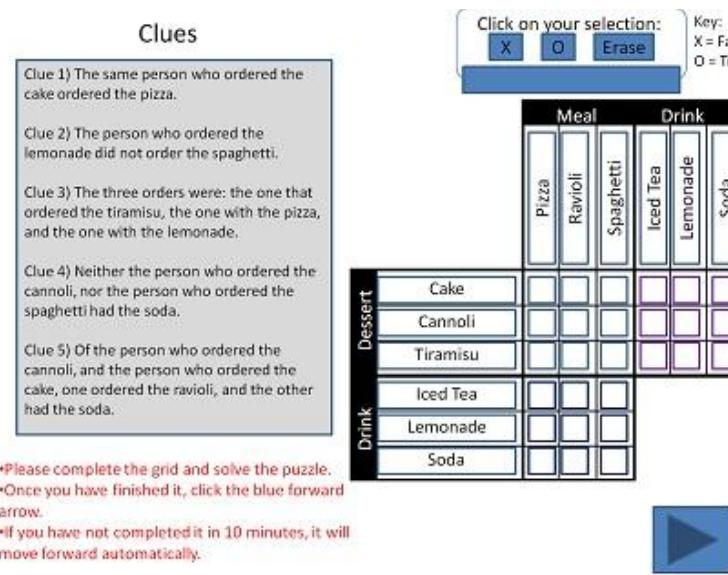
Figure 2. A screenshot example of a logic grid puzzle that is included as domain content in GIFT 4.0.

Before we review recent work on non-cognitive personalization in the context of the CT, MATHia, and logic grid puzzles, we summarize previous research on personalized content for learning. We then provide several recommendations for GIFT and important questions for future research.

Related Research

Personalization Research

Context personalization can be described as adjusting learning material to align with the interests of an individual learner (Anand & Ross, 1987; Cordova & Lepper, 1996; Walkington, 2013). While the general concept of personalizing material has been consistent, approaches to examining personalized material in the literature have varied. Among the varying techniques that have been studied are the impact of providing examples and learning materials that are consistent with one's own major (Ross, 1983; Ross, McCormick & Krisak, 1986), using self-rated user interests to determine the topics of examples that are provided (Walkington, 2013), providing instruction written in the first person (Moreno & Mayer, 2000), and using self-generated names or interests within the learning materials (Anand & Ross, 1987; Cordova & Lepper, 1996; Ku & Sullivan, 2002; Sinatra, 2013). The research approaches generally fall into two categories: (1) changing the topics of the learning material to something that the individual states that they find interesting and (2) encouraging learners to relate the information to themselves. Despite different approaches to personalization, the results have been fairly consistent. Personalizing content does appear to have a positive effect on learning outcomes, primarily in transfer performance, or taking the learned material one step further and applying it in a new situation. One explanation for this effect may be that relating information in the learning material to the individual's interests, or the self, leads to better understanding of the material, and deeper learning. This deeper learning is then demonstrated by the individual through the ability to work with and apply the learned information in a new way.

There are at least two approaches to context personalization: at the individual level or at the classroom level. In traditional classrooms and examples, it may be easier to find topic areas that many of the students are familiar with and use these to personalize instruction (Ku & Sullivan, 2002; Ross, 1983).

This is an easy task for an instructor and may show some improvements due to interest and motivation. However, in these cases, there may be individual students who do not share the interests and will not receive the benefits. With the increase in availability of computerized learning and ITSs, we now have the ability to individually adapt the context of materials for each student. Research has shown positive outcomes from adapting material to include student entered information (Anand & Ross, 1987; Cordova & Lepper, 1996; Sinatra, 2013), as well as student-selected topic areas (Ross et al., 1986; Walkington, 2013).

Why Personalization Works

Two types of explanations for why personalization improves learning are common; the first is affective and the second is cognitive. Affective explanations often suggest that personalization increases the likelihood that the learner will be inherently interested in the material, resulting in higher motivation to engage with it. Engagement with the material leads to the individual paying closer attention to it and feeling enjoyment while completing the lessons or assessments. Findings have been consistent that individuals report enjoying personalized material more than non-personalized material (Anand & Ross, 1987; Cordova & Lepper, 1996; Ross, 1983; Ross et al., 1986). An additional affective explanation is at the system level. Learners may be more attentive or more conscientious within a system that attends to and acknowledges personal factors. This could lead to positive feelings toward the system, as the learner feels acknowledged. In Cordova and Lepper (1996), the reactions of elementary school students support this interpretation, as some offered exclamations of joy when seeing their names present in the learning materials. These positive feelings may enhance engagement with the material during the learning phase, leading to better performance and transfer performance.

Cognitive explanations often suggest that linking learned information to something that the learner already understands reduces mental workload. By providing this context, particularly one of interest, the individual expends less mental energy understanding the question itself and can spend more cognitive resources on understanding it. Such explanations are consistent with the Cognitive Theory of Multimedia Learning (Mayer, 2005), which suggests that relieving cognitive load frees up other resources to engage with the material. For example, a student is familiar with baseball would have an easier time understanding a math word problem that explains a baseball game rather than an unfamiliar area, such as a chemistry experiment. Specifically, Ross (1983) found better learning outcomes for nursing students who received medical examples (as opposed to education examples) and for teachers who received education examples (as opposed to medical examples). It was suggested that having an interest and knowledge in the subject matter increased motivation (Ross, 1983; Ross et al., 1986). An additional explanation posits that students already had a foundation in the area, which assisted in providing context and reducing mental workload. Further, this foundation may lead to better comprehension of the materials and problems that are being presented to the learner (Anand & Ross, 1987). If fewer resources are used to make sense of the problem, more resources will be available to learn from it and successfully solve it. The self-reference effect, or benefit of linking information to one's self (e.g., by including one's own name in the materials), may reduce cognitive workload in a similar way to context personalization. However, there is no consensus in the literature regarding which specific cognitive mechanisms are responsible for the self-reference effect (Klein, Loftus & Burton, 1989; Symons & Johnson, 1997).

Discussion

Recent research using the Cognitive Tutor ITS considers the effect of personalizing content of mathematics word problems based on student interest areas outside of the classroom. Experimental studies (e.g., Walkington 2013) have determined that tailoring word problems to domains like sports or shopping tends

to improve student performance on parts of problems that involve especially easy and difficult KCs; the effect for more difficult KCs is greater when the readability of the word problem's text is especially high-level.

In contrast to fine-grained analysis of student performance on opportunities to practice specific (groups of) KCs, recent analysis of data from MATHia (Fancsali & Ritter, 2014) does not find a strong association between the extent to which this product "honors" student preferences (as measured by the proportion of problems that honor students' favored interest areas) and aggregate indicators of student performance and learning. This lack of association may be due to students' relatively low frequency of receiving personalized problems in this naturalistic environment. However, this analysis did provide a novel finding that might point to a systemic affective influence: students who express "strong" interest area preferences (e.g., rating "sports & fitness" with the maximum rating while rating all other areas with the minimum rating) tend to perform better than those students who set "weak" preferences (e.g., rating all interest areas the same), and students who merely express interest area preferences tend to perform better than students that do not. Importantly, in both comparisons, students tend to spend roughly the same amount of time logged into MATHia, so setting of such preferences does not function as a mere proxy for this sort of engagement. Students who set names of classmates or friends for inclusion in the text of word problems also tend to perform better in MATHia. These findings are only correlational and not experimental, but they do suggest that students who are engaged by the personalization factors tend to get better outcomes. Further analyses of these findings are an important topic for future research. Specifically, differences between students who set classmates' names are interesting in light of recent research on similar types of content personalization we now briefly summarize.

The majority of personalization research focuses on adjusting problem context to be consistent with the interests of an individual. However, when personalization takes the form of providing links to the self, its effects may arise due to mechanisms similar to those at play for the self-reference effect, which has consistently shown that making information relevant to the self can improve recall (Symons & Johnson, 1997). Moreno and Mayer (2000) found that by providing science lesson material in the first-person rather than the third-person led to improved learning transfer performance. Further, d'Ailly, Simpson, and MacKinnon (1997) found that including the word "you" had similar positive effects in mathematics word problems. One of the proposed reasons for this effect was that receiving lessons that include the word "you" encouraged learners to think of themselves while encoding the information. The learning materials in both Anand and Ross (1987) and Cordova and Lepper (1996) included student-provided information such as own name and the names of close friends. However, in these studies, personalization also included favorite topics and other interests. Therefore, it is not clear what type of personalization (the self or interests) drives the learning benefits.

A small pilot study (Sinatra, 2013) using GIFT examined the impact of name personalization as part of the content of an interactive logic grid puzzle tutorial. The puzzle within the tutorial either included self-provided names of the individual and friends, names of popular culture characters, or names that were not expected to have meaning to the individual. See Figure 3 for an example of the popular culture and general name conditions of the logic grid puzzle tutorial. The goal of the manipulation was to examine if including names that activate a schema for the self or a familiar set of characters leads to positive learning outcomes from a tutorial. Results from the pilot were inconclusive; however, initial results from the full study have suggested that individual differences, such as the need for cognition (NFC) (Cacioppo & Petty, 1982), may have an impact on the benefits of personalization. Name personalization/self-reference and NFC (how much an individual enjoys thinking) did not impact general assessment questions regarding the content of the tutoring. However, name personalization/self-reference and NFC did have an impact on transfer performance (the individual's ability to solve a puzzle that was more difficult than the one included during the tutorial). Those who were high in NFC had significantly better transfer performance when they received self-personalized materials than those who were low in NFC. These results

suggest that there may be individual characteristics that can impact whether or not personalization is a helpful strategy for assisting learning. As these initial results were found with a less well-defined area, future studies should examine if NFC has an impact on personalization in more traditional domains such as math and science.

Figure 3. Screenshot examples of the name manipulation from the logic grid puzzle tutorial used in Sinatra (2013). The popular culture manipulation is on the left, and the generic manipulation is on the right.

Future personalization research should closely examine the type of personalized materials that are used (interests or self) and determine which are most beneficial to include in instruction. In addition, it is important to begin examining individual differences that may interact with personalization and consider the influence of simply asking personalization questions, independent of what use is made of this information. While the systemic affective results may point to a positive effect resulting from asking personalization questions, research in the area of stereotype threat suggests that priming individuals by asking them to think about their gender or ethnicity prior to test performance can activate negative stereotypes and hurt their performance (Wheeler & Petty, 2001). Thus, the attempt to personalize based on gender or ethnicity has the potential for producing a negative effect. Further research is needed to fully understand the factors that cause personalization surveys to have positive or negative effects.

The bulk of personalization research has been done in the area of mathematics and arithmetic (Anand & Ross, 1987; Cordova & Lepper, 1996; d'Ailly et al., 1997; Ku & Sullivan, 2002; Ross, 1983; Ross et al., 1986; Walkington, 2013), and researchers would benefit from expanding our knowledge of the influence of personalization in other areas. As personalization appears to be most beneficial to deeper learning and transfer performance, it is important to start branching out into new areas that are less well defined and determine if consistent results are found.

Recommendations and Future Research

Given results that indicate positive effects for different types of content personalization in ITSs, it is important that GIFT provide a flexible framework that allows for different types and levels of personalization (e.g., such that researchers and instructors have the ability to set the extent to which a tutor "honors" different kinds of student preferences) and for malleable, possibly evolving, student preference settings. GIFT provides a survey mechanism whereby students could set, for example, interest areas, but it is unclear whether and how a student might change such settings, for example, over the course of a semester. In addition, previously stated concerns about the influence of surveying itself on outcomes

suggests that GIFT would benefit from a wide variety of options in collecting such information. In some cases, we anticipate that system designers will want to highlight the connection between the survey and the system action (e.g., "I selected this problem for you because you said you liked sports."), where in other cases (perhaps with ethnicity or gender), we might not want system adaptation to be linked to the underlying personalization factor.

Ever-present, survey-like mechanisms could help to constantly refine the tutor's knowledge of student preferences. Students, for example, could "rate" particular problems they are presented (e.g., something as simple as a "thumbs up" when a student likes a problem) to allow for iterative improvement of content personalization. Such persistent survey-like mechanisms may not be well represented in the GIFT architecture, as they are dissimilar, for example, in some ways from traditional physical sensors for which the GIFT architecture makes allowances. Currently, GIFT can select content based on prior surveys, but it does not have a mechanism for adapting content (e.g., the ability to insert student-entered names into authored content). It would be beneficial for GIFT to provide an option to authors to include student-entered content within authored questions. Further, communication among GIFT modules should be sufficiently rich to allow for adaptive presentation of personalized course content based on a multitude of factors. GIFT should also provide mechanisms for customizing the transition between problems (or other types of content). This kind of facility would allow researchers and instructional designers to make the factors that led to selection of particular content either more or less evident. Enhanced flexibility for personalization in GIFT will provide opportunities to increase the learner's enjoyment and comprehension of material.

References

Anand, P.G. & Ross, S.M. (1987). Using computer-assisted instruction to personalize arithmetic for elementary school children. *Journal of Educational Psychology, 79* (1), 72 – 78.

Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., & Rossi, L. (2012). Sensor-free automated detection of affect in a Cognitive Tutor for algebra. In *Proceedings of the 5th International Conference on Educational Data Mining (EDM 2012)*, 126 – 133.

Cacioppo, J.T. & Petty, R.E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42* (1), 116 – 131.

Conati, C., Chabbal, R. & Maclaren, H. (2003). A study on using biometric sensors for detecting user emotions in educational games. *Workshop on "Assessing and Adapting to User Attitude and Affects: Why, When, and How?"* in conjunction with User Modeling 2003, 9th International Conference on User Modeling, Pittsburgh, USA.

Conati, C. & Maclaren, H. (2004). Evaluating a probabilistic model of student affect. In *Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS 2004)*, 55 – 66.

Cordova, D.I. & Lepper, M.R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88* (4), 715 – 730.

d'Ailly, H. H., Simpson, J. & MacKinnon, G.E. (1997). Where should "You" go in a math compare problem? *Journal of Educational Psychology, 89* (3), 562 – 567.

Fancsali, S.E., & Ritter, S. (2014). Context personalization, preferences, and performance in an intelligent tutoring system for middle school mathematics. Submitted to *4th International Conference on Learning Analytics and Knowledge,* 73-77.

Fancsali, S.E., Ritter, S., Stamper, J., & Nixon T. (2013). Toward "hyper-personalized" Cognitive Tutors. In *AIED 2013 Workshops Proceedings Volume 7*, 71 – 79.

Farrington, C.A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T.S., Johnson, D.W. & Beechum, N.O. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review.* Chicago: University of Chicago Consortium on Chicago School Research.

Klein, S.B., Loftus, J. & Burton, H.A. (1989). Two self-reference effects: The importance of distinguishing between self-descriptiveness judgments and autobiographical retrieval in self-referent encoding. *Journal of Personality and Social Psychology, 56* (6), 853 – 865.

Koedinger, K.R., McLaughlin, E.A. & Stamper, J.C. (2012). Automated student model improvement. In *Proceedings of the 5th International Conference on Educational Data Mining (EDM 2012)*, 17 – 24.

Ku, H.Y. & Sullivan, H.J. (2002). Student performance and attitudes using personalized mathematics instruction. *Educational Technology Research and Development, 50*(1), 21 – 34.

Mayer, R.E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, 31 – 48.

Moreno, R. & Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Educational Psychology, 92* (4), 724 – 733.

Pane, J.F., Griffin, B.A., McCaffrey, D.F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale. *Educational Evaluation and Policy Analysis*, 36 (2), 127-144.

Ritter, S., Anderson, J.R., Koedinger, K.R. & Corbett, A.T. (2007). Cognitive Tutor: applied research in mathematics education. *Psychonomic Bulletin and Review*, 14 (2), 249 – 255.

Ross, S.M. (1983). Increasing the meaningfulness of quantitative material by adapting context to student background. *Journal of Educational Psychology, 75* (4), 519 – 529.

Ross, S.M., McCormick, D. & Krisak, N. (1986). Adapting the thematic context of mathematical problems to student interests: Individualized versus group-based strategies. *Journal of Educational Research, 79*, 245 – 252.

Sinatra, A. M. (2013). Using GIFT to support an empirical study on the impact of the self-reference effect on learning. In *AIED 2013 Workshops Proceedings Volume 7*, 80 – 87.

Sottilare, R. A. & Holden, H. K. (2013). Motivations for a Generalized Intelligent Framework for Tutoring (GIFT) for authoring, instruction and analysis. In *AIED 2013 Workshops Proceedings Volume 7*, 1 – 9.

Sottilare, R.A. & Proctor, M. (2012). Passively classifying student mood and performance within intelligent tutors. *Educational Technology & Society*, 15 (2), 101 – 114.

Symons, C.S. & Johnson, B.T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin, 121* (3), 371 – 394.

Walkington, C.A. (2013). Using adaptive learning technologies to personalize instruction to student interests: The impact of relevant contexts on performance learning outcomes. *Journal of Educational Psychology*, Advance online publication, 1 – 14.

Wheeler, S.C. & Petty, R.E. (2001). The effect of stereotype activation on behavior: A review of possible mechanisms. Psychological Bulletin, 127 (6), 797 – 826.

CHAPTER 7 – Adaptive Interventions to Address Students' Negative Activating and Deactivating Emotions during Learning Activities

Ivon Arroyo[1], Kasia Muldner[2], Winslow Burleson[3], Beverly P. Woolf[4]
[1] Worcester Polytechnic Institute; [2] Arizona State University; [3] New York University;
[4] University of Massachusetts Amherst

Introduction

Students' emotions can positively or negatively influence their achievement outcomes: confidence, boredom, confusion, stress, and anxiety are all strong predictors of achievement (Goleman, 1996; Pekrun, Goetz, Daniels, Stupinsky & Perry, 2010) and affective predispositions such as low self-concept and pessimism diminish academic success (Corno & Snow, 1986; Kluger & DeNisi, 1998; Seligman, 1991; Sweller, Van Merriënboer & Paas, 1998). As far as science, technology, engineering, and mathematics (STEM) topics are concerned, females, minorities, and students with learning disability experience more frustration and anxiety when solving problems than do their peers (Arroyo, Mehranian & Woolf, 2010; Frenzel, Pekrun & Goetz, 2007; Woolf et al., 2010). It is not surprising that these students also anticipate more barriers in STEM activities and more bias in their self-assessments (Correll, 2001; McWhirter, 1997). Thus, it is critical to understand how to address negative emotions as it occurs for each student. However, while students learn best with personalized instruction (Jonassen & Grabowski, 1993; Rose & Meyer, 2002), it is difficult to provide such instruction in standard classrooms that typically have one teacher for every 20 or more students.

Teachers do attend to the affective needs of individual students (Lepper & Hodell, 1989; Rosiek, 2003), but they have very limited means to recognize and respond to students' affect in a typical classroom. Given the reality of already burdened teachers and school systems, the dream of individualized education may be turned into reality only through adaptive tutoring technologies that supplement traditional classroom instruction. It is not surprising that interest has emerged in developing affect-aware technologies, given the pivotal role that affect and motivation play in the success of learning activities. The overwhelming majority of this work to date, however, has focused on *modeling* affect, i.e., designing computational models that infer how students feel while interacting with an intelligent tutoring system (ITS) (Arroyo et al., 2009; Conati & Maclaren, 2009; Cooper et al., 2009; Cooper et al., 2010; D'Mello & Graesser, in press; D'Mello & Graesser, 2007; Muldner, Burleson & VanLehn, 2010). While modeling affect is a critical first step in providing adaptive support tailored to students' affective needs, very little work exists to systematically explore the impact of affective interventions on students' performance, learning, affect, and attitudes, i.e., how to respond to students' emotions, such as frustration, anxiety, boredom, and hopelessness, as they arise.

The work described here takes steps to fill this gap through a series of proposed interventions that address student emotion in computer-based learning environments. We focus on repairing negative valence student emotions, as these can be especially detrimental to student learning (Pekrun et al., 2010). We propose ways to approach the automatic repair of negative affective states in digital learning environments, once the emotion is recognized, and provide an example of one successful intervention that has already shown positive results at repairing negative states of boredom and low excitement while learning.

The testbed for our work is Wayang Outpost[1], a computational tutor for mathematics that teaches effectively (Arroyo et al., 2007; Arroyo, Woolf, Royer & Tai, 2009; Arroyo et al., 2009). Tens of thousands of students in the United States have already used this tutoring software and have consistently shown significant learning gains, e.g., on mathematics tests (an increase of 12% from pre- to post-test after only 4 class periods), on state standard exams (92%) as compared to students not using Wayang (76%), and on Measures of Academic Progress (MAP), a national test provided by Northwestern Education Association (NWEA).

Background And Theoretical Perspectives

We base our discussion and our prior research on the control-value (CV) theory of achievement emotion in education (Pekrun, 2006). This theory describes emotions related to achievement in learning situations that may be classified according to students' value and valence (positive/negative), their focus (within activity/prospective/retrospective), and perceived level of control (high/low). We add that these emotions can be activating or deactivating (high/low arousal). *Positive activating* achievement emotions (enjoyment of activity, hope/confidence) have a positive impact on achievement (Pekrun, Goetz, Titz & Perry, 2002), while *negative deactivating* emotions (boredom and hopelessness) and *negative activating* emotions (anxiety) have a negative impact (Pekrun et al., 2010; Zeidner, 2007).

Over the last five years, we have developed software that detects a subset of emotions defined in the CV theory. We chose four specific emotions, namely confidence (a bipolar scale equivalent to the control-value theory's hope and anxiety, prospective emotions), interest (a bipolar scale equivalent to CV's engaged concentration vs. boredom, activity-based emotions), frustration (a unipolar scale equivalent to CV's frustration and anger, activity-based emotions), and excitement (a unipolar scale equivalent to CV's enjoyment, activity-based emotion). In the past, we devised computational models that recognized these four emotions in real time, using a variety of data sources, and found that the models achieved high accuracy (>78%) and reasonable kappa values – for details, see Arroyo et al., 2009; Cooper et al., 2009; Shanabrook, Arroyo, Woolf & Burleson, 2012). Once emotions are detected, the system can use this information to respond appropriately to address students' affective needs.

Within the perspective of the CV theory of achievement emotions, a key element in design of affective interventions is that students experience specific achievement emotions when they feel *in/out* of *control* of activities and outcomes that are subjectively important to them. The theory is also dependent upon how students *value* the subject (e.g., how important it is to know mathematics), and how *capable* they perceive themselves (e.g., expectation of success). For instance, if students feel that their failures are due to lack of innate ability and not a natural part of learning, they will tend to perceive more *anxiety* in class (Weiner, 2007). This implies that *control appraisals* and *value appraisals* are proximal determinants of these emotions, and that helping students properly attribute success/failure to their activities, helping them feel in control of their learning and to think intrinsically that the topic is valuable to learn (Schutz & Pekrun, 2007) can help students modify the likelihood that certain emotions will occur, and to transition into other more positive emotions.

Addressing Student Emotional States

We suggest that interventions to address and repair student emotion should: (1) target student beliefs about the self and the task (value-oriented interventions), (2) target student's self-regulation strategies to help students (self) regulate their emotions and their learning process in effective ways (control-oriented

---

[1] Wayang is freely available at http://wayangoutpost.com. Click on "guest user login."

interventions) and/or (3) manipulate the learning context (context-oriented interventions) to keep students within the zone of proximal development (ZPD) (Murray & Arroyo, 2002; Vygotsky, 1978), acknowledging that the context (difficult or repetitive tasks) can influence emotions too.

These interventions can be broadly classified as providing three kinds of support and interventions: *affective*, *cognitive/contextual*, and *metacognitive,* and all are aimed at reducing either activating emotions (frustration and anxiety) or negative deactivating emotions (boredom and apathy), see Table 1. We believe that these two sets of emotions should be dealt with in different ways. The ultimate goal is to develop the full suite of interventions shown in Table 2. We expect this suite of interventions to increase students' positive feelings about themselves, their interactions with others, their perspective about mathematics classes, and learning. Here, we describe steps we have taken thus far in these directions.

Table 1. Three kinds of support to repair negative activating/deactivating affective states.

Student Emotion	Affective Support (Value Oriented)	Cognitive Support (Context Oriented)	Metacognitive Support (Control Oriented)
Frustration Anxiety Negative Activating	1. Acknowledgement (visual/verbal mirroring) of emotion, then attribution training	2. Maintain student in the Zone (decrease problem difficulty) Offer or force hints and worked-out examples (increase cognitive support)	3. Train active resolution strategies (seeking help, making sketches, simplifying problem, decoding givens and unknowns)
Boredom Disinterest Negative Deactivating	4. Train math value (show videos of experts using math in other areas, e.g., architecture)	5. Maintain students in the Zone (increase difficulty). Peer scaffolding (invite collaboration)	6. Support self-reflection, then goal setting, then give control via choice

Table 2. Sample attribution training and growth mindset training messages used by learning companions.

Type	Sample Message
Attribution (General)	"I found out that people have myths about math, thinking that only some people are good in math. The truth is we can all be good in math if we try."
Attribution (Effort)	"Keep in mind that when we are struggling with a new skill we are learning and becoming smarter!"
Attribution (No Effort)	"We will learn new skills only if we are persistent. If we are very stuck, let's call the teacher, or ask for a hint!"
Attribution (Incorrect)	"When we realize we don't know why the answer was wrong, it helps us understand better what we need to practice."
Effort Affirmation (Correct No-effort)	"That was too easy for you. Let's hope the next one is more challenging so that we can learn something."
Effort Affirmation (Correct Effort)	"Good job! See how taking your time to work through these questions can make you

	get the right answer?"
Strategic (Incorrect)	"Let's think again: What are the steps we have to carry out to solve this one?"
Strategic (Correct)	"We are making progress; can you think of what we have learned in the last problems?"

The next section describes how we envision adding a variety of scaffolds to address a student's emotion including *negative activating emotion* and *negative deactivating emotion*. We have studied some scaffolding and interventions for these emotions that are discussed within the context of the Wayang Tutors, used by students in grades 6–10.

Negative Activating Emotion

Student emotions can be activating or deactivating (high/low arousal). Positive activating achievement emotions, including interests, enjoyment, hope/confidence, have a positive impact on achievement (Pekrun et al., 2002). However, negative activating emotions, including frustration and anxiety, have a negative impact on student learning and behavior. We have applied three interventions *affective*, *cognitive/contextual*, and *metacognitive* to reduce negative activating emotions, see Table 1.

Providing Affective Support for Negative Activating Emotions (Table 1, Support #1)

In our past work, we integrated gendered learning companions (Figure 1) into Wayang Outpost, which offered advice and encouragement by talking to students, see Arroyo, Burleson, Tai, Muldner & Woolf (2013). These companions can gesture and train attributions for "success/failure" that have been associated with positive learning outcomes, e.g., intelligence is malleable, perseverance and practice are needed to learn, making mistakes is an essential part of learning, and failure is not due to a lack of innate ability (see sample messages, Table 2). The learning companions have demonstrated their utility for fostering positive student affect. Specifically, in controlled randomized studies with hundreds of students, certain groups of students (females and students with disabilities) reported decreased frustration and increased confidence levels when working with learning companions and increased frustration when companions were not present (Arroyo et al., 2013). In addition, student enjoyment, self-efficacy, and interest were higher compared to students not given learning companions, suggesting that such affective pedagogical agents can impact students' emotions (Arroyo, Woolf, et al., 2010; Woolf et al., 2010).
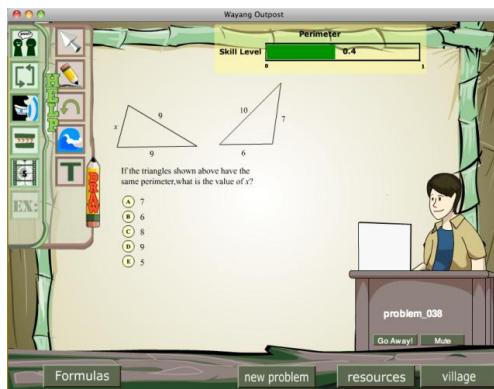
Figure 1. Learning companions use gestures to offer advice and encouragement. Students can ask for hints or click the "solve it" button. Animations, videos, and worked-out examples add to the spoken hints about steps in a mathematics problem.

However, to date, our own learning companions have not responded to learners' emotional states and have acted in a "preventive" manner, regardless of student emotion. As a result, and despite positive significant effects for the overall population of students, characters seemed to have been "harmful" to one group of students (high achieving males), who had higher affective baselines at pretest time; the characters seem to have been distracters for this group of students (Arroyo, Woolf, et al., 2010; Woolf et al., 2010). Characters were more effective for the lower achieving students in the class (lower than median math performance at pretest time) and for female students in general. These results suggest that affective characters should probably not be provided and/or be refined for students who are not presently frustrated or anxious, or otherwise, and that high achieving male students might need a different form of support (which has yet to be determined).

Accordingly, as part of the affective (value-oriented) support, we are now working on the design of characters that will respond directly to students' emotions. It is possible that this will be an optimal solution for all students, with characters speaking up as negative emotions arise, in a two-phase process: (1) empathize with a student's emotional reaction (e.g., "I understand, as sometimes I also get [frustrated] when solving math problems") and (2) resolve the situation by training failure attributions and growth mindset, (e.g., "However, struggling in problems is actually a good thing, because it means that we are learning something new and becoming wiser").

Providing Cognitive Support for Negative Activating Emotions (Table 1, Support #2)

The *zone of proximal development* (ZPD) has been defined as a region of ideal challenge, where a student is neither bored nor frustrated (Csikszentmihalyi, 1990; Murray & Arroyo, 2002; Vygotsky, 1978). One goal of personalized tutors is to guide students into their own Zone and into a state of "flow," a level of problem complexity at which students can succeed with effort, probably with some help from an expert, which is the computer in this case. When students are outside of that Zone, they are either confused and frustrated, or bored and disinterested (Murray & Arroyo, 2002).

We propose that positive (more pleasurable) emotions are experienced when students are within the ZPD, where they are working within a possible level of success, and that negative (less pleasurable) emotions are experienced above and below the Zone, see Figure 2. Movement happens along the Zone from bottom left to top right, as students progress in learning. For example, a student might begin a project and be curious to understand the problem, but not know much about the topic, see (1) in Figure 2. The student may solve "easy" problems correctly and thus remain in the Zone, confident about his or her problem-solving ability. However, if the student is provided challenging problems at this initial phase and his or her solutions are wrong several times, the student might move into a space of frustration or anxiety, see (3) in Figure 2. Meanwhile, if the material is too easy or repetitive, the student may fall into the area of boredom, see (2) in Figure 2 (Pekrun et al., 2010). Thus, a typical learning experience involves a range of emotions and one goal of the field of intelligent tutoring systems is to modify material within the tutoring to keep students within the Zone. Part of this involves increasing/decreasing the difficulty of learning activities as needed and providing extra support when students are in negative activating space, see (3) in Figure 2, and providing worked-out examples and scaffolds when students face difficulties, to propel them back into the Zone again, see (1) in Figure 2.

Early work defined the ZPD as the interplay of students' skill level and the difficulty of the material presented, using entirely ad-hoc parameters that defined whether a student was inside or outside of that Zone. This previous work depended entirely on the student's cognitive ability. A tutoring system that is additionally aware of students' emotions can help create a fuller model and a better understanding of whether the student is inside or outside of the Zone. For example, detecting confusion (a prior stage to frustration – both are activating and have negative valence) provides evidence that a student is moving outside of the ZPD. Decreasing the level of challenge and providing extra support seems logical in this

case. However, increasing the challenge might not be sufficient when a student falls into negative deactivating states (see (2) in Figure 2), e.g., boredom.
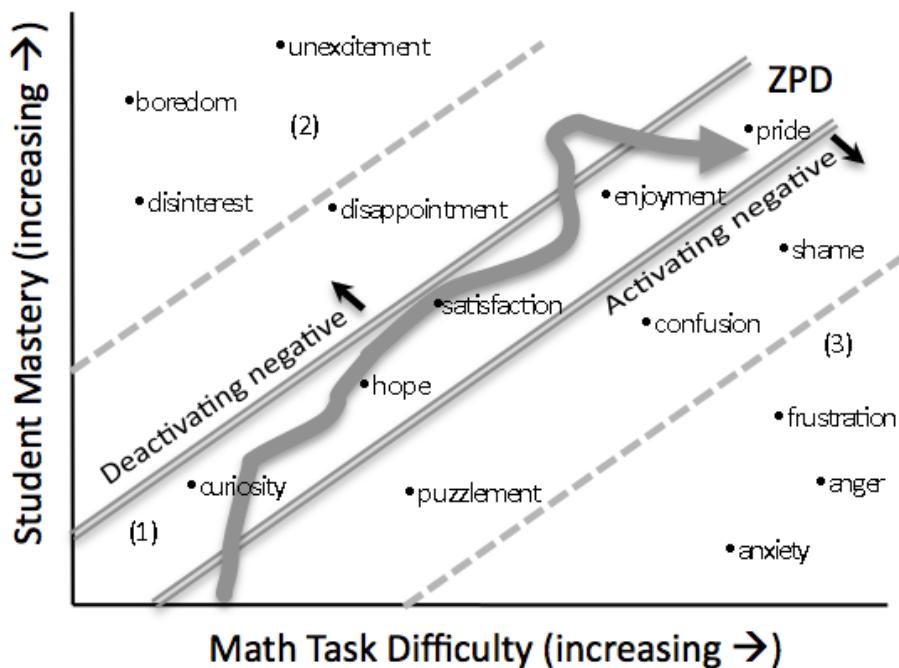
Figure 2. Keeping students within the ZPD (large meandering grey arrow), where positive valence achievement emotions are generally experienced, is important for learning. Expressed emotions help to clarify whether students are inside or outside of that Zone. Area (1) represents an ideal Zone, (2) is a zone of *negative deactivating* emotions, and (3) a zone of *negative activating* emotions.

Providing Metacognitive Support for Negative Activating Emotions (Table 1, Support #3)

Metacognition, the knowledge about when and how to use strategies for learning, e.g., self-regulation and executive control, has the potential to address the *control* component of the CV theory of emotion. Students use a variety of suboptimal coping strategies to regulate their emotions in stressful learning situations, including humor and acceptance, social-emotional coping, abandoning/avoidance, and negation (Eynde, de Corte & Verschaffel, 2007), suggesting they need better and more productive strategies to cope. Some work has been done to address metacognitive aspects of learning; e.g., one tutoring system tried to correct unproductive behavior such as students either avoiding hints or rushing through hints (Aleven, Roll, McLaren, Ryu & Koedinger, 2005). However, the results were not encouraging, probably because the tutor was too reactive – it stopped other possible interactions when students displayed inappropriate behaviors and suggested that their behavior was not productive, possibly reinforcing their frustration and making students feel increased lack of control (which, according to the CV theory, was the origin of the negative emotion in the first place).

We propose a modification to this approach, in which interventions give control back to students. When students experience *negative activating* emotions (frustration, anxiety), the learning companions will deliver metacognitive tips to seek out further interventions, "This is a good moment to ask for a hint, how about clicking on help." In a proactive manner, companions will not judge students' behavior, rather will suggest on-the-spot strategies to encourage students to persevere in the problem, e.g., "How about we use the pencil tool to make a sketch to help us solve this one," if there is no figure accompanying the problem,

84

or "Hey, how about we click on the Read Aloud button," if the tutor inferred that the student rushed to an action, without enough time to read the problem.

Negative Deactivating Emotion

Negative deactivating emotions, including boredom, hopelessness, and apathy, have a negative impact on learning and behavior (Pekrun et al., 2010; Zeidner, 2007). We propose three types of support and interventions: *affective*, *cognitive/contextual*, and *metacognitive* aimed at reducing negative deactivating emotions, see Table 1.

Providing Affective Support for Negative Deactivating Emotions (Table 1, Support #4)

We assumed that the affective support proposed for frustration and anxiety would improve negative activating emotions. According to the CV theory of emotions, attribution training should reduce shame, anger, and anxiety. We did show this for anxiety (Arroyo et al., 2013), especially for those students with highest affective needs at pretest time. Also, interest increased (and boredom reduced) on average, thanks to the support provided by the affective learning companions provided. However, as we conjectured, the larger effect sizes were for frustration and anxiety, i.e., negative activating emotions. Specifically frustration was a full standard deviation lower for girls who received the female learning companion as compared to girls who did not receive any companion.

We still believe that other approaches are needed when dealing with negative deactivating emotions, even when talking about affective and value-oriented scaffolds. For instance, we think the *value* and *usefulness* of mathematics in life should be trained directly to recover interest and enjoyment activity-based emotions. We propose to show the value of *videos of domain experts* highlighting the importance and applications of mathematics topics addressed at the moment, in concrete settings.

Providing Cognitive Support for Negative Deactivating Emotions (Table 1, Support #5)

D'Mello and colleagues have called the state of boredom the "deadly loop," or a state that is difficult for a student to leave (Baker, D'Mello, Rodrigo & Graesser, 2010; Graesser et al., 2006) – suggesting that a radical change is needed when students experience boredom. We propose *collaborative learning* as a novel approach to address boredom and to bring students back into the Zone. In general, collaborative activities have been successful in mathematics classrooms, exhibiting large effect sizes for achievement in standardized test scores as compared to control groups (Johnson & Johnson, 1999; Slavin, Lake & Groff, 2009). Part of this success has been accompanied by enhanced affective and motivational outcomes, improved attitudes by people who collaborate (Johnson, Johnson, Johnson & Anderson, 1976), such as more altruism and positive attitudes toward classroom life. Others have found that collaboration increased self-esteem, social acceptance, and peer ratings, particularly for students with disabilities (Putnam, Markovchick, Johnson & Johnson, 1996). We think that inviting collaboration can help address negative states such as boredom because it provides a *break in modality* and adds *social value*. Intelligent learning environments that promote collaborative learning for mathematics have been scarce (but notable exceptions exist, e.g., Stahl, 2006; Walker, Rummel & Koedinger, 2011).

We will explore the utility of collaborative activities once boredom and disinterest have been detected. We have begun implementing synchronous face-to-face cooperative activities that involve coordinated problem solving with *neighboring students*. Wayang Outpost has a record of which students are sitting next to which student (e.g., at login time, students identify neighbors) and so can invite students to work together in a shared activity when negative deactivating emotions emerge. In this way, the tutor may choose which common activities to assign to pairs of student depending on each student's proficiency and

emotional profile, and provide shared support (e.g., examples, tutorials, etc.) to the pair. Models of student mastery for each student as well as group models of how pairs work together will mediate tutoring actions in upcoming work.

This implies that there will be, in general, two choices for collaboration for a bored student: the student to the left and the student to the right, each of which have their own emotional states, cognitive profile, gender, friends, etc. In addition, the activity proposed to students could affect the pair – is it a challenging activity for both? Is it a challenging activity for one but an easy activity for the other? These are not minor details, as the collaboration between two bored students might not work (it might reinforce boredom/disinterest), while the collaboration of an interested student with a bored student could bring the bored student back to an interested state. To increase the likelihood of pairs of students working well together, we are considering having friends sit next to each other, since this appears to be beneficial (Azmitia & Montgomery, 1993).

We have piloted this idea of "casual collaboration" with students in grade 7–8 (Arroyo, Woolf & Shanabrook, 2012). In that pilot experiment, students were given a special "Go to" button that allowed them to move to a specified problem – in this way a student and his or her neighbor could both "go to" the same problem and work on it together. We then identified students who collaborated on a problem in this manner, and analyzed engagement levels for students in both collaborative moments vs. "solo" moments. Collaboration with neighboring students on the same math problem increased students' incidence of engagement behaviors and decreased incidence of disengagement behaviors such as gaming the system (Arroyo et al., 2012). The fact that collaboration helped to change student behavior from gaming to productive behavior (e.g., hints requests, time on task) suggests that inviting students to collaborate can be a valuable way to at least repair student disengagement.

Providing Metacognitive Support for Negative Deactivating Emotions (Table 1, Support #6)

When negative deactivating emotions emerge (e.g., boredom), the situation is complicated by the fact that rehearsal is positively correlated with boredom, suggesting that repeating the same mode of activities could lead to boredom (Pekrun et al., 2010); thus, a more radical change is needed. We have evidence already that a progress page, or learning dashboard, can improve negative deactivating emotions, as a consequence of helping students self-regulate: students feel less "lost" in the learning process, are encouraged to set goals, and can self-reflect on progress toward those goals. A variety of positive deactivating emotions accompany an enhanced self-regulatory learning experience, such as enhanced interest, reduced boredom, and increased enjoyment of the learning process. Our hypothesis aligns with Zimmerman & Moylan's (2009) model of self-regulation that integrates metacognitive processes and key measures of motivation, during three phases: forethought, performance, and self-reflection. According to this model, learning dashboards support forethought and self-reflection (former and latter phases), while motivation and affect are extremely important in the self-reflection and forethought phases.

We support the self-regulatory cycle with a student progress page, which is offered at key moments when negative deactivating emotions occur, by asking students to (1) reflect on their knowledge; (2) select a guided choice over a variety of goals and topics; and (3) choose to continue/review/challenge or switch to another topic. We based this intervention on past positive results that provided basic progress charts that helped to repair disengagement and also yielded higher learning rates (Arroyo et al., 2007).

To date, we have devised, implemented and analyzed the impact of a *Student Progress Page* (SPP) that encourages students to stop to think (a metacognitive process) when they become disinterested or unexcited. The algorithm consists of the following steps: (1) when low interest or low excitement is detected, offer students a chance to reflect on their knowledge: "Would you like to see how you are doing?" (2) upon acceptance, take students to a SPP (Figure 3) where they can reflect on the state of their knowledge

and effort (plants bloom and produce fruits upon effort instead of mastery), while the mastery bar focuses on demonstrated math ability and progress); (3) provide reflections about how they are doing for each topic, and provide a guided choice of topic and actions ("continue," "review," or "challenge me"); and 4) go back to the tutor in this new "mode".
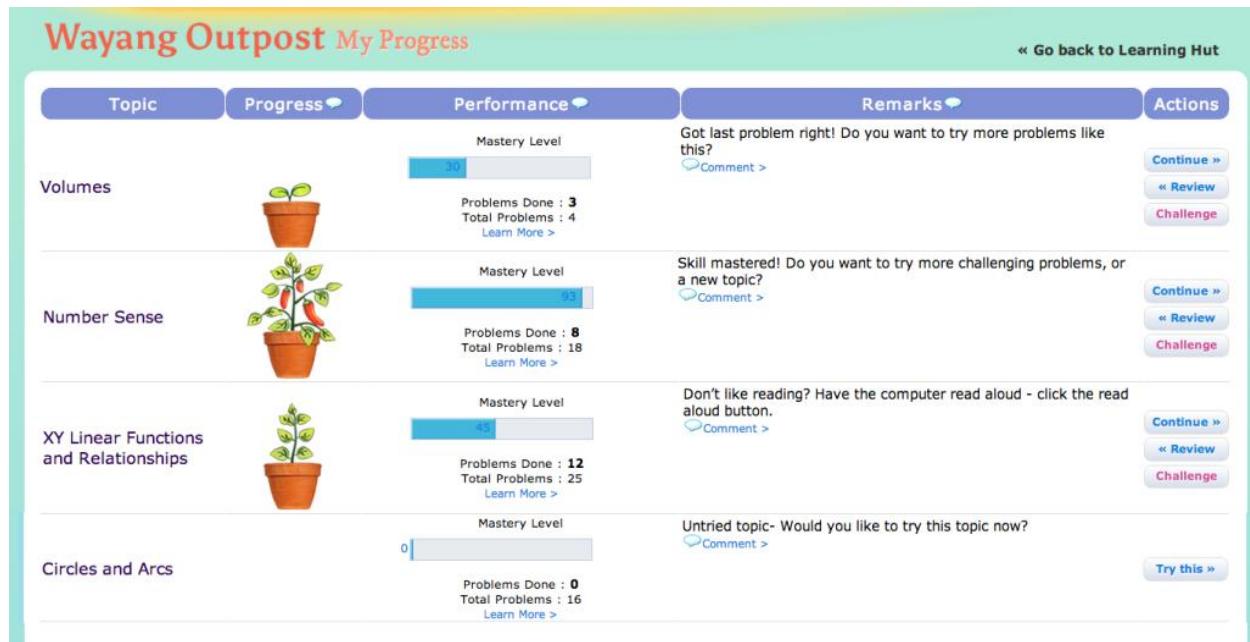
Figure 3. The SPP encourages students to reflect on their effort (plants, column 2) for each math topic, reflect about their mastery (bars, column 3), reflect about their recent behaviors (column 4), make informed decisions about reviewing problems they got wrong, or challenging themselves with harder problems (column 5).

We have evaluated this support in three studies, with positive preliminary results. In a more recent study, students in an experimental condition were invited to use the learning dashboard when the tutor detected a negative deactivating state (boredom or lack of excitement), compared to students in the control who had the same learning dashboard available via a button, but were not offered the dashboard at those key moments. Unfortunately, students were almost never bored in this experiment, so there were almost never offered the SPP. This resulted in two groups of students who were basically identical in amount of total accesses to the student progress page.

We thus changed the analysis and considered students who had more vs. less frequent accesses to the Student Progress Page, splitting students on the median number of total SPP accesses. The results based on this split showed that students with high accesses to such metacognitive scaffolding experienced higher enjoyment and interest. More detailed analyses, shown in Figure 4, showed that students were more likely to transition from a neutral emotional state to states of high interest (+0.22 more likely), instead of remaining in a neutral state or becoming bored. This is demonstrated by a Markov chain analysis, a path model that shows how students transition between emotional states.
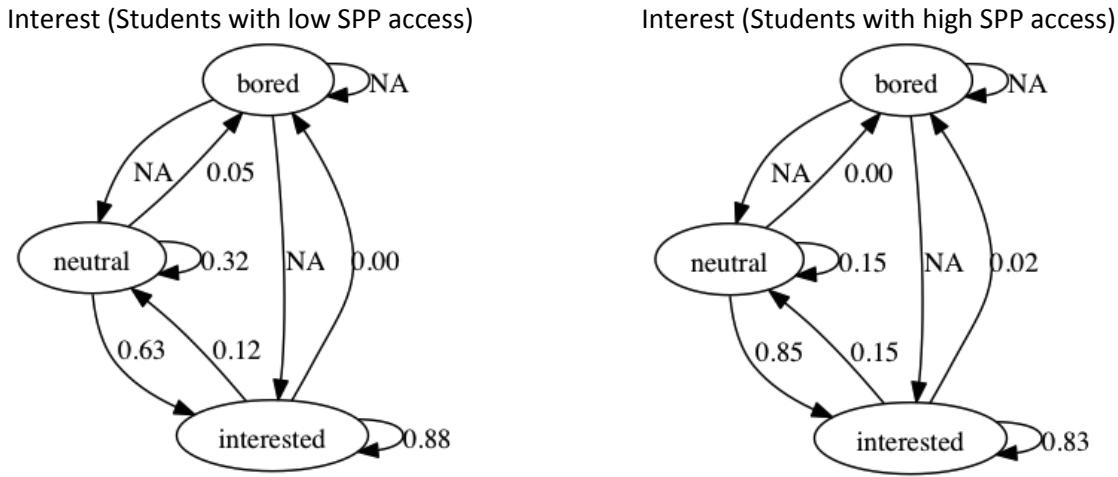
Figure 4. Student transition probabilities from one learning activity to another, between states of boredom and interest. The overall likelihood for students remaining interested was 83% for students who accessed the SPP most (right, bottom), a bit lower than the 88% for students in the control condition (left, bottom). However, students with high SPP access had a higher likelihood to transition from neutral to interested (0.85) than did students in the control condition (0.63). Students with High SPP access were less likely to remain in the neutral state (0.15) than students with lower SPP access (0.32).

Recommendations and Future Research

We have outlined a variety of approaches to address activating vs. deactivating emotions, starting off from the CV theory of achievement emotions, by addressing the root elements of control and value, in addition to cognitive factors such as challenge level, and level of scaffolding. We provided experimental evidence that a few of these approaches do work at repairing negative emotions.

A couple of issues should be addressed. The first one is the need to identify the different ways to respond to activating vs. deactivating emotions, with the assumption that negative activating emotions (e.g., frustration) require more immediate and localized repairs. On the other hand, negative deactivating emotions (e.g., boredom) might need a more global approach, including scaffolding, which makes the learning experience more valuable. Such scaffolding might train students to stop and think about the learning.

Another issue is that both a *preventive* and a *repair* approach might be needed to address student affect. In fact, we mentioned how characters that attempted to train a growth mindset instilled positive affective states, even if those messages were not particularly delivered at key moments of negative emotions, at least for female and low-achieving students. On the other hand, a student progress page that targeted self-regulation and was proposed to the student at key moments of boredom and lack of excitement, was better than a control condition in which students were not brought to the page, but still had it available.

Another issue is our assumption that disengagement, affect, and motivation are different but overlapping constructs at different levels, and that benefits in one actually lead to benefits at other levels. We based some of our results on previous successes that focused on disengagement; motivational approaches such

as training attributions for failure/success; and growth mindset come from the motivation line of research, yet, as we have shown, they have an impact at the emotional experience level. In general, we consider that any intervention that makes the learning experience more positive, including cognitive-oriented interventions (e.g., better ways to provide support, better understanding of the interface and tools available) and metacognitive interventions (e.g., teaching students to self-regulate their learning process), should also benefit students at an affective level.

Future research consists of evaluating all the proposed ways to enhance student affective states, and more importantly, conducting a larger study where each intervention competes with each other to see which ones are more effective at which moments, and for which specific emotions.

Acknowledgment

References

Aleven, V., Roll, I., McLaren, B., Ryu, E.J. & Koedinger, K. (2005). An architecture to combine meta-cognitive and cognitive tutoring: Pilot testing the Help Tutor. *Artificial Intelligence in Education: Supporting Learning Through Intelligent And Socially Informed Technology, 125*, 17.

Arroyo, I., Burleson, W., Tai, M., Muldner, K. & Woolf, B. (2013). Gender Differences In the Use and Benefit of Advanced Learning Technologies for Mathematics. *Journal of Educational Psychology*, 105 (4), 957 - 969.

Arroyo, I., Ferguson, K., Johns, J., Dragon, T., Meheranian, H., Fisher, D.,Barto, A., Mahadevan, S., Woolf, B.P... (2007). *Repairing disengagement with non-invasive intervention.* Paper presented at the 13th International Conference on Artificial Intelligence in Education, Marina del Rey, CA.

Arroyo, I., Mehranian, H. & Woolf, B. (2010). *Effort-based tutoring: An empirical approach to intelligent tutoring.* Paper presented at the 3rd International Conference on Educational Data Mining, Pittsburgh, PA.

Arroyo, I., Woolf, B.P., Royer, J.M. & Tai, M. (2009). *Affective gendered learning companions.* Paper presented at the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling.

Arroyo, I., Woolf, B.P., Royer, J.M., Tai, M., Muldner, K., Burleson, W. & Cooper, D. (2010). Gender Matters: The Impact of Animated Agents on Students' Affect, Behavior and Learning *Technical report UM-CS-2010-020*: UMASS Amherst.

Arroyo, I., Cooper, D.G., Burleson, W., Woolf, B.P., Muldner, K. & Christopherson, R. (2009). *Emotion Sensors Go To School.* Paper presented at the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling.

Arroyo, I., Woolf, B.P. & Shanabrook, D. (2012). *Casual Collaborations while Learning Mathematics with an Intelligent Tutoring System.* Paper presented at the Intelligent Tutoring Systems 2012: Workshop on Support for Exploratory Environments.

Azmitia, M. & Montgomery, R. (1993). Role of friendship in peer discussion of scientific reasoning. *Social Development, 2*(3), 202-221.

Baker, R.S.J.d., D'Mello, S.K., Rodrigo, M.M.T. & Graesser, A.C. (2010). Better to Be Frustrated than Bored: The Incidence, Persistence, and Impact of Learners' Cognitive-Affective States during Interactions with Three Different Computer-Based Learning Environments. *International Journal of Human-Computer Studies, 68*(4), 223-241.

Conati, C. & Maclaren, H. (2009). *Modeling User Affect from Causes and Effects.* Paper presented at the User Modeling and User Adapted Interaction (UMAP).

Cooper, D., Arroyo, I., Woolf, B., Muldner, K., Burleson, W. & Christopherson, R. (2009). *Sensors Model Student Self Concept in the Classroom.* Paper presented at the UMAP 2009, First and Seventeenth International Conference on User Modeling, Adaptation and Personalization, Trento, Italy.

Cooper, D.G., Muldner, K., Arroyo, I., Woolf, B.P., Burleson, W. & Dolan, R. (2010). *Ranking Feature Sets for Emotion Models used in Classroom Based Intelligent Tutoring Systems.* Paper presented at the International Conference on User Modeling and Adaptive Presentation (UMAP'10), Hawaii.

Corno, L. & Snow, R. (1986). Adapting teaching to individual differences among learners. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 605-629). New York: Macmillan.

Correll, S. J. (2001). Gender and the Career Choice Process: The Role of Biased Self-Assessments. *The American Journal of Sociology, 106*(6), 1691-1730.

Csikszentmihalyi, M. (1990). *Flow : the psychology of optimal experience* (1st ed.). New York: Harper & Row.

D'Mello, S. & Graesser, A. (in press). Language and Discourse are Powerful Signals of Student Emotions during Tutoring. *IEEE Transactions on Learning Technologies*. doi: 10.1109/tlt.2012.10

D'Mello, S. & Graesser, A. (2007). *Mind and Body: Dialogue and Posture for Affect Detection in Learning Environments.* Paper presented at the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work.

Eynde, P., de Corte, E. & Verschaffel, L. (2007). Student Emotions: A key component of self-regulated learning? *Emotion in education*. San Diego, CA: Academic Press.

Frenzel, A. C., Pekrun, R. & Goetz, T. (2007). Girls and mathematics—a "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 22*(4), 497-514.

Goleman, D. (1996). *Emotional Intelligence: why it can matter more than IQ*. London.

Graesser, A.C., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S. & Gholson, B. (2006). *Detection of Emotions During Learning with AutoTutor.* Paper presented at the 28th Annual Conference of the Cognitive Science Society.

Johnson, D., Johnson, R., Johnson, J. & Anderson, D. (1976). Effects of Cooperative Versus Individualized Instruction on Student Prosocial Behavior, Attitudes Toward Learning and Achievement. *Journal of Educational Psychology, 68*, 446-452.

Johnson, D. W. & Johnson, R. (1999). *Learning together and alone: Cooperative, competitive, and individualistic learning* (First edition, 1975 ed.). Boston: Allyn & Bacon.

Jonassen, D. H. & Grabowski, B. L. (1993). *Handbook of Individual Differences: Learning and Instruction*: Lawrence Erlbaum.

Kluger, A. N. & DeNisi, A. (1998). Feedback Interventions: Toward the Understanding of a Double-Edged Sword. *Current Directions in Psychological Science,, 7*, 67-72.

Lepper, M. R. & Hodell, M. (1989). Intrinsic motivation in the classroom. In C. Ames & R. E. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 73-105). New York: Academic Press.

McWhirter, E. (1997). Perceived Barriers to Education and Career: Ethnic and Gender Differences. *Journal of Vocational Behavior, 50*, 124-140.

Muldner, K., Burleson, Winslow & VanLehn, Kurt. (2010). *"Yes!": Using Tutor and Sensor Data to Predict Moments of Delight during Instructional Activities.* Paper presented at the User Modeling, Adaptation, and Personalization.

Murray, T. & Arroyo, I. (2002). *Towards measuring and maintaining the zone of proximal development in adaptive instructional systems.* Paper presented at the 6th International Conference on Intelligent Tutoring Systems.

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review., 18*, 315–341.

Pekrun, R., Goetz, T., Daniels, L., Stupinsky, R. & Perry, R. (2010). Boredom in Achievement Settings: Exploring Control–Value Antecedents and Performance Outcomes of a Neglected Emotion. *Journal of Educational Psychology, 102*(3), 531-549.

Pekrun, R., Goetz, T., Titz, W. & Perry, R. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist, 37*(2), 91-105.

Putnam, J., Markovchick, K., Johnson, D. W. & Johnson, R. T. (1996). Cooperative learning and peer acceptance of students with learning disabilities. *The Journal of Social Psychology, 136*, 741-752.

Rose, D.H. & Meyer, A. (2002). *Teaching Every Student in the Digital Age: Universal Design for Learning*. Alexandria, VA: Association for Supervision & Curriculum Development.

Rosiek, J (2003). Emotional Scaffolding: An Exploration of Teacher Knowledge at the Intersection of Student Emotion and the Subject Matter. *Journal of Teacher Education,, 54*(5), 399-412.

Schutz, P. A. & Pekrun, R. (2007). *Emotion in education*. San Diego, CA.: Academic Press.

Seligman, M. (1991). *Learned Optimism.* New York: Knopf.

Shanabrook, D.H., Arroyo, I., Woolf, B.P. & Burleson, W. (2012). *Visualization of Student Activity Patterns within Intelligent Tutoring Systems.* Paper presented at the Intelligent Tutoring Systems (ITS'2012).

Slavin, R., Lake, C. & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research, 79*(2), 839-911.

Stahl, G. (2006). Sustaining group cognition in a math chat environment. *Research and Practice in Technology Enhanced Learning (RPTEL), 1*(2), 85-113.

Sweller, J., Van Merriënboer, J. & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251-296.

Vygotsky, L. (1978). *Mind in society*: Harvard University Press.

Walker, E., Rummel, N. & Koedinger, K. R. (2011). Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer-Supported Collaborative Learning, 6*(2), 279-306.

Weiner, B. (2007). Examining emotional diversity in the classroom: an attribution theorist considers the moral emotions. In Schutz PA & Pekrun R (Eds.), *Emotions in education* (pp. 75–88). San Diego: Academic Press.

Woolf, B., Arroyo, I., Muldner, K., Burleson, W., Cooper, D., Dolan, R. & Christopherson, R. (2010, June 2010). *The Effect of Motivational Learning Companions on Low Achieving Students and Students with Disabilities.* Paper presented at the The 10th International Conference on Intelligent Tutoring Systems (ITS'10), Pittsburgh, PA.

Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 165-184). San Diego, CA: Academic Press.

Zimmerman, B. J. & Moylan, A. R. (2009). Self-regulation: Where motivation and metacognition intersect *Handbook of Metacognition in Education*.

CHAPTER 8 – Assessing Persistence in Educational Games

Matthew Ventura[1], Valerie Shute[1], and Matthew Small[2]
[1]Florida State University; [2]Independent researcher

Introduction

In this chapter, we describe how assessing persistence in educational games can be useful in enhancing learning. We describe Newton's Playground (NP), a 2D video game aimed at helping students understand qualitative physics. The goal across all of the 74 puzzles in NP is for the player to guide a green ball to a red balloon. The primary way to move the ball is by drawing on the screen with a mouse. Once objects are drawn they are affected by gravity and Newton's three laws of motion. We describe results from a study where we mined data from NP log files to develop an in-game measure of persistence and report on how persistence in NP predicts learning gains. Finally, we propose various ways to use the detection of persistence to enhance learning in educational games using the GIFT framework.

There is growing evidence of video games and simulations supporting learning (e.g., Coller & Scott, 2009; Tobias & Fletcher, 2011; for a review, see Wilson et al., 2009). Playing digital games has been shown to be positively related to a variety of cognitive skills (e.g., visual-spatial abilities, Green & Bavelier, 2007; Ventura, Shute, Wright & Zhao, 2013; attention, Shaw, Grayson & Lewis, 2005), personality traits (e.g., Openness, Chory & Goodboy, 2011; Ventura, Shute & Kim, 2012; Witt, Massman & Jackson, 2011), persistence (Ventura, Shute & Zhao, 2012), academic performance (e.g., Skoric, Teo & Neo, 2009; Ventura, Shute, Kim, 2012), and civic engagement (Ferguson & Garza, 2011). Moreover, educational games have been shown to enhance learning of academic content, within and outside of the game (e.g., Barab, Dodge et al., 2010; Coller & Scott, 2009; DeRouin-Jessen, 2008).

In addition to video games' effects on learning, they produce a vast amount of data that can be used for assessment purposes (Dede, 2005; DiCerbo & Behrens, 2012; Quellmalz, Timms, Silberglitt & Buckley 2012). Using this stream of data, formative assessments that are embedded in games can enable us to more accurately provide feedback and change gameplay to maximize learning according to the ability level of the player. One way to meet these requirements is to use *stealth assessment* (Shute, 2011). Stealth assessment refers to assessments that are woven directly and invisibly into the fabric of the gaming environment. During game play, students naturally produce rich sequences of actions while performing complex tasks, drawing on the very skills or competencies that we want to assess (e.g., scientific inquiry skills, creativity). Evidence needed to assess the skills is thus provided by the players' interactions with the game itself, which can be contrasted with a typically singular outcome of an activity – the norm in educational environments.

Making use of this stream of gameplay evidence to assess students' knowledge, skills, and understanding (as well as beliefs, feelings, and other learner states and traits) presents problems for traditional measurement models used in assessment. First, in traditional tests the answer to each question is seen as an independent data point. In contrast, the individual actions within a sequence of interactions in a game are often highly dependent on one another. For example, what one does in a particular game at one point in time affects subsequent actions later on. By analyzing a sequence of actions within a quest (where each response or action provides incremental evidence about the current mastery of a specific fact, concept, or skill), stealth assessments within game environments can infer what learners know and do not know at any point in time.

The main assumptions underlying educational games include the following: (a) learning by doing (required in game play) improves learning processes and outcomes, (b) different types of learning and

learner attributes may be verified and measured during game play, (c) strengths and weaknesses of the learner may be capitalized on and bolstered, respectively, to improve learning, and (d) ongoing feedback can be used to further support student learning. In line with these assumptions, a central question in this chapter is, *Can we enhance learning in educational games through the detection of persistence?* Persistence (i.e., industriousness in Roberts, Chernyshenko, Stark & Goldberg, 2005; achievement in Perry, Hunter, Witt, Harris, 2010) is a facet of conscientiousness that reflects a dispositional need complete difficult tasks (McClelland, 1961) and the desire to exhibit high standards of performance in the face of frustration (Dudley, Orvis, Lebiecki & Cortina, 2006; Ventura, Shute & Zhao, 2012). Over the past 20 years or so, conscientiousness has emerged as one of the most important personality traits in predicting academic performance (e.g., Poropat, 2009) as well as in various life outcomes (e.g., Roberts, Kuncel, Shiner, Caspi & Goldberg, 2007). Perry et al. (2010) suggest that persistence may drive the predictive validity of conscientiousness and is the facet that consistently predicts a variety of outcomes (Dudley et al., 2006; Perry et al., 2010; Roberts et al., 2005) over other facets of conscientiousness.

Persistence can play an important role in learning in a video game due to the design principle of challenge in well-designed games (Pausch, Gold, Skelly & Thiel, 1994). That is, in games, establishing the right level of challenge entails adjusting the optimal level of difficulty for a player. This balance – between game difficulty and player ability level – is consistent with the theory of the ZPD (Vygotsky, 1978), which states that learning takes place right at the outer edges of one's abilities. The principle of challenge is pervasively used in video games and has been shown to engage attention and enhance learning (Lepper & Malone, 1987; Rieber, 1996; Sweetser & Wyeth, 2005). Thus video games can require persistence due to the design of progressive difficulty. This repeated exposure to challenge can positively affect persistence requiring a willingness to work hard despite repeated failure (for a review, see Eisenberger, 1992; Ventura, Shute & Zhao, 2012).

To illustrate this idea of repeated exposure to challenge and the relationship to persistence, Eisenberger and Leonard (1980) showed that trying to complete difficult tasks can improve persistence. Participants were randomly assigned to solve impossible, hard, or easy anagrams and then take the perceptual comparison task where they were asked to detect as many differences as possible between two pictures. Participants who had experienced the impossible anagram condition spent the most time on the perceptual comparison task, followed by those in the hard anagram condition, and then those in the easy anagram condition. This provides evidence that exposure to difficult tasks can affect effort on subsequent tasks. The next section introduces a video game we developed that requires persistence due to its difficulty.

Qualitative Physics in Newton's Playground

Research into what's called "folk" physics demonstrates that many adults hold erroneous views about basic physical principles that govern the motions of objects in the world, a world in which people act and behave quite successfully (Reiner, Proffit & Salthouse, 2005). The prevalence of these systematic errors has led some investigators to propose that incorrect performance on these tasks is due to specific "naive" beliefs, rather than to a general inability to reason about mechanical systems (McCloskey & Kohl, 1983). Recognition of the problem has led to interest in the mechanisms by which physics students make the transition from folk physics to more formal physics understanding (diSessa, 1982) and to the possibility of using video games to assist in the learning process (Masson, Bub & Lalonde, 2011; White, 1994).

One way to help remove misconceptions in physics is to illustrate physics principles with physical machines (Hewitt, 2009). In physics, a machine refers to a device that is designed to either change the magnitude or the direction of a force. Teaching about simple machines (e.g., lever, pulley, and wedge) is widely used as a method to introduce physics concepts (Hewitt, 2009). Research on science education

also indicates that learners' hands-on experience with such machines (both virtually and physically) support applicable understanding of important physics concepts (Hake, 1998).

We developed a computer video game called Newton's Playground (Shute & Ventura, 2013) to help middle school students better understand *qualitative physics* (Ploetzner & VanLehn, 1997). Qualitative physics is a nonverbal understanding of Newton's three laws, balance, mass, gravity, conservation of momentum, potential, and kinetic energy. NP is a nonlinear video game (i.e., a game design feature that allows many possible solution paths) where players are challenged to guide a green ball to a red balloon in a 2D environment, where all objects obey the basic rules of physics relating to gravity and Newton's three laws of motion. The player can add a slight impulse to the ball by clicking on it, but the primary way to guide the ball is by drawing simple machines on the screen that "come to life" once the mouse button is released (a gameplay mechanic inspired from games such as Magic Pen and Crayon Physics Deluxe). Using this drawing mechanic, players can create any shape imaginable to help move the ball toward the goal and solve the level. Players can also use pins, which act as a rotating joint, to attach objects to one another.

The 74 problems in NP require the application of four categories of simple machines: inclined plane/ramps, levers, pendulums, and springboards. These simple machines, which we often refer to as "agents of force and motion," or just "agents," are created by a combination of the player drawing colored lines on the screen in various shapes (i.e., objects) and attaching objects to each other using pins. A ramp is an object that helps to guide a ball in motion. It can be useful to transform vertical motion to horizontal motion (and vice versa) or guide the ball over a hole. A lever rotates around a fixed fulcrum or pivot point and is generally useful when a player wants to move the ball vertically. The rotation of the lever can be generated by the weight distribution of the lever itself or by an external object transferring momentum to the lever on one side of the fulcrum. A pendulum swings on a pin and directs an impulse tangent to its direction of motion. With enough space, a pendulum can be used to exert a horizontal or vertical force. A springboard (or diving board) stores elastic potential energy provided by a falling weight. Springboards provide an efficient mechanism to move the ball vertically.

Figure 1 displays a puzzle in NP. In this puzzle, the player must draw a pendulum on a pin (i.e., little black circle) to make it swing down to hit the ball (surrounded by a heavy container hanging from a rope). In the depicted solution, the player drew a pendulum that will swing down to move the ball. To succeed, the player should manipulate the mass distribution of the club and the angle from which it was dropped to accomplish just the right amount of force to get the ball to the balloon.
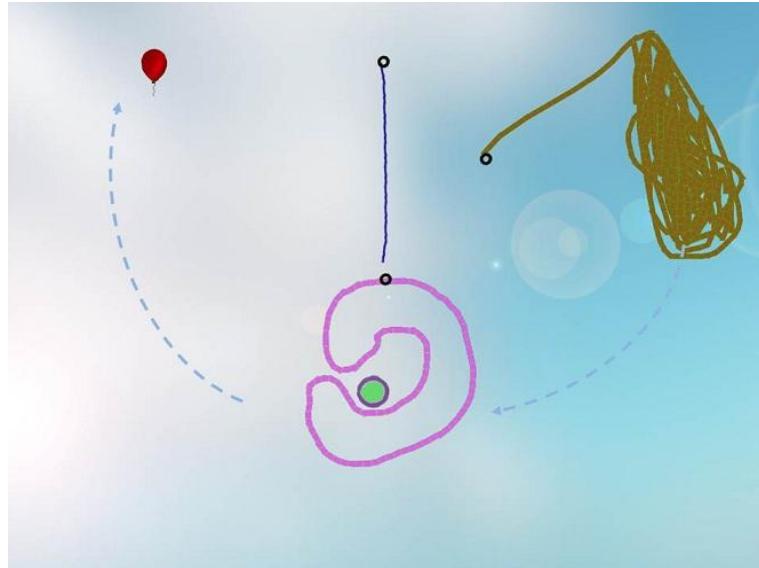
Figure 1. Example problem in NP.

Other Gameplay Features

NP consists of 7 playgrounds (each one containing 10–11 problems) that progressively get more difficult. Each problem is designed to elicit a particular set of simple machines (in the game we refer to them as "agents"). The difficulty of a problem is based on a number of factors including: relative location of ball to balloon, obstacles, number of agents required to solve the problem, and novelty of the problem. NP also includes tutorial videos that show the player how to create and use the various agents. During gameplay, students have the option to watch agent tutorial videos at any time.

NP displays silver and gold trophies in the top left part of the screen, which represent progress in the game. A silver trophy is obtained for any solution to a problem. Players can also receive a gold trophy if a solution is under a certain number of objects (the threshold varies by problem, but is typically < 3). A player can receive one silver and one gold trophy per problem.

NP Session Logs

NP automatically uploads log files to a server for each gaming session (i.e., log activity between login and logout). The text below displays what a session log looks like for one *event* of a puzzle. An event collects data for a particular visit to a puzzle. A player may revisit a puzzle multiple times thus logging multiple events. Figure 2 displays a snapshot of the NP session event log. As can be seen, the session event log reports several features of gameplay in a puzzle. For example, "game_time" reports the total time (in seconds) spent on this particular visit to the problem. "Silver" reports if a silver trophy was achieved in this visit to the problem.

```
"time_stamp" : 12.163,
"level_path" : ".\\levels\\p4\\diving board.level",
"game_time" : 130.526001,
"pause_time" : 1.54,
"restart_count" : 7,
"object_count" : 14,
"object_limit_count" : 2,
"nudge_count" : 42,
"erase_count" : 13,
"pin_count" : 9,
"agent_vector" :"61.78 SB, 98.08 SB, 131.60 SB",
"ball_trajectory" : "<0.733, 0.427> <0.766, 0.394>.
"silver" : true,
"gold" : false,
"solved" : true
```

Figure 2. NP session event log.

The Impact of Persistence on Learning Gains in Newton's Playground

Ventura and Shute (2013) analyzed log files from 70 8th and 9th grade students who played NP for around four hours (split into five 45-minute sessions across two weeks) in a large computer lab. Based on the theory of the persistence (Ventura, Shute, Zhao, 2012), we developed a game-based assessment of persistence (GAP), which is derived from time spent on *unsolved problems* over all events in the player's log file over the five sessions. That is, longer times spent on difficult problems (whether they were solved or not) should indicate greater persistence (Eisenberger & Leonard, 1980; Ventura, Shute & Zhao, 2012). Time on solved problems should not be an indicator of persistence since solution times are primarily based on skill in the game. The time spent on each unsolved problem was summed across all events from the log file over the five sessions. For example, if a player attempted (but did not solve) a problem 10 different times, the time spent on that problem would be summed across all 10 attempts. The average time is then taken for all the unsolved problem sums (out of a possible 74 problems).

As part of our validation process, we also administered another performance based measure of persistence consisting of impossible anagrams and picture comparison tasks (see Ventura & Shute, 2013). Impossible anagrams consist of jumbled letters that do not actually make a word. Impossible picture comparison items consist of two adjacent pictures where participants are told to detect difference between pictures when in fact no differences exist. At any time the individual can also choose to select the "skip" button to leave the current trial and go on to the next one. If the individual guesses correctly, the person is told that he or she is correct, and is presented with a new trial. The score from the anagrams and picture comparison tests is the time spent on impossible trials since these times represent effort expended on frustrating tasks.

Ventura and Shute (2013) found that the GAP (i.e., unsolved times) predicted a number of learning gains for struggling players in NP. Table 1 displays the correlations among the GAP and other measures of persistence and learning. As can be seen, the GAP significantly relates to the anagrams and picture comparison tasks and the physics post-test scores. Additionally, the GAP relates to the post-test scores of low performers even after controlling for gender, video game experience, physics pretest, and enjoyment ($pr = 0.26$, $p < 0.05$) suggesting that both persistence measures predict learning even after controlling for background knowledge and game enjoyment.

Table 1. Correlations among performance measures in NP (Ventura & Shute, 2013).

	A-PC	GAP	Gold
GAP	0.47[**]		
Gold	0.00	0.14	
Physics post-test	0.30[*]	0.31[*]	0.08

* = p < 0.05; ** = p < 0.01; A-PC = anagrams and picture comparison task; GAP = unsolved times

The results from Ventura and Shute (2013) suggest that persistence can positively impact learning for struggling students. The GAP was positively correlated with the posttest even after controlling for gender, video game experience, pretest knowledge, and enjoyment of NP. This suggests that the GAP plays an important role in learning. Additionally, the GAP was correlated with the anagrams and picture comparison tasks, another measure of persistence. Thus we found evidence of construct validity of the GAP. Both of these persistence measures are grounded on the premise that longer times spent on difficult problems indicate persistence (Eisenberger & Leonard, 1980; Ventura, Shute & Zhao, 2012).

The usefulness of the GAP does appear to depend on whether kids were being sufficiently challenged in the game. That is, players who had more difficulty in the game were operating under the required conditions to elicit persistence (i.e., to persist one must be challenged). This is consistent with the theoretical framework of the persistence which requires students to expend effort on really hard or impossible problems.

Enhance Learning in Games with GAP

Despite repeated claims that persistence is a highly valuable skill needed for success in school, on the job, and in life in general (e.g., Roberts, Kuncel, Shiner, Caspi & Goldberg, 2007), there is no prior empirical research testing the relationship between persistence and learning in educational games. Ventura and Shute (2013), however, provide preliminary evidence that a relation exists between persistence and learning. We believe that future work should explore this relationship further, as well as other ways to enhance learning in games by integrating GAP into educational game design. Next, we discuss three potential methods to apply GAP data in educational game design: tuning gameplay difficulty, hints, and feedback.

A primary goal of using GAP to inform the tuning of gameplay difficulty is to keep players in the ZPD, where they are challenged enough to exhibit persistence (and experience its potential benefits) but are not so challenged that they are unable to complete any game objectives or experience excessive frustration. The tuning itself can be performed nearly continuously during gameplay (e.g., adjustments to the quantity and strength of generated enemies) or at a larger quantum (e.g., serving more difficult levels). Optimally selecting the granularity of tuning is largely dependent on game structure and the granularity of GAP data points. We plan to develop additional fine grained data points for NP that inform GAP calculations more continuously (e.g., in-game drawing data), but these techniques will likely be more specific to the game mechanics of NP and thus less transferable to other games.

Hints are another potential target for exploiting GAP within game design to enhance learning. Hints can provide an effective resource for guiding players through difficult, unfamiliar, or poorly understood game scenarios. A constraint to the overall efficacy of hint generating systems (e.g., ITSs) is access to informative data points about the current state of the player's skills and conceptualizations. Since a primary goal of a hint generating system is to provide players with minimal guidance to complete game objectives (i.e., keep players in the ZPD), GAP could prove an effective tool in producing more targeted hints.

We are also interested in the potential benefits of providing players with explicit and real-time feedback about their GAP level. By providing players with such information regarding their persistence, we are essentially providing the player with a new metric by which to measure their gameplay performance, like a badge or trophy. This metric could become a tool for players to self-regulate their persistence and potentially learn to become more persistent by doing so. Designing GAP feedback would need to be done with care to both give players some indication of potential strategies for regulating their GAP and avoid exposing so much about GAP calculations as to allow "gaming of the system."

Persistence in the Generalized Intelligent Framework for Tutoring

We have outlined a method to detect persistence demonstrated within educational games. One advantage of the simplicity of the data needed to detect persistence is that it can be applied to a variety of educational products. The main aim of GIFT is to support varying open and dynamic game-based learning environments that apply distinctively different messaging protocols. This involves embedding components and processes within GIFT's domain module to support the detection of persistence regardless of the educational game being utilized.

GIFT requires rules and models built around game interaction that must be explicitly linked to concepts defined inside of the GIFT architecture. For this purpose, a 'Gateway Module' is incorporated that associates an external educational/training system's state data with a domain or competency model built within GIFT. This linkage allows for two disparate systems to communicate with one another. In the case of persistence within GIFT, this enables the application of real-time assessment of persistence in players.

This allows any system to link interaction with GIFT's domain model, where persistence assessments are conducted and progress is communicated to the learner model for determining transitions in performance or competency. This approach to assessment is ideal in game-based environments as tracking interaction data as it relates to objectives can denote comprehension and understanding that is difficult to gauge in traditional assessment techniques. The application of stealth assessment within GIFT potentially provides further diagnosis of game performance, which can be communicated to the pedagogical model for more focused selection of feedback and remediation tactics.

References

Coller, B. D. & Scott, M. J. (2009). Effectiveness of using a video game to teach a course in mechanical engineering. *Computers & Education, 53*(3), 900-912.

Dede, C. (2005). Planning for neomillennial learning styles. *EDUCAUSE Quarterly, 28*(1), 7–12.

DiCerbo, K. E. & Behrens, J. T. (2012). Implications of the digital ocean on current and future assessment. In R. Lissitz & H. Jiao (Eds.) *Computers and their impact on state assessment: Recent history and predictions for the futur*e (pp. 273–306). Charlotte, NC : Information Age Publishing.

diSessa, A. A. (1982). Unlearning Aristotelian physics: A study of knowledge-based learning. *Cognitive Science, 6,* 37–75.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006).A meta-analytic investigation of Conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91,* 40–57.

Eisenberger, R. (1992). Learned industriousness. *Psychological Review, 99,* 248-267.

Eisenberger, R. & Leonard, J. M. (1980). Effects of conceptual task difficulty on generalized persistence. *American Journal of Psychology, 93*, 285-298.

Ferguson, C. J. & Garza, A. (2011). Call of (civic) Duty: Action games and civic behavior in a large sample of youth. *Computers in Human Behavior, 27,* 770–775.

Hake, R. R. (1998). Interactive engagement vs. traditional methods in mechanics instruction. *American Journal of Physics, 66*(1), 64–74.

Hewitt, P. G. (2009). *Conceptual physics* (11th ed.). San Francisco, CA: Pearson Education.

Lepper, M. R. & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. In R. E. Snow & M. J. Farr (Eds.), *Conative and Affective Process Analysis* (pp. 255-286). Hillsdale: Lawrence Erlbaum Associates, Inc.

Masson, M. E. J., Bub, D. N. & Lalonde, C. E. (2011). Video-game training and naive reasoning about object motion. *Applied Cognitive Psychology, 25*, 166–173.

McCloskey, M. & Kohl, D. (1983). Naive physics: The curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 9, 146–156.

Pausch, R., Gold, R., Skelly, T. & Thiel, D. (1994). What HCI designers can learn from video game designers. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 177-178). ACM Press.

Ploetzner, R. & VanLehn, K**.** (1997). The acquisition of informal physics knowledge during formal physics training. *Cognition and Instruction*, 15(2), 169-205.

Perry, S. J., Hunter, E. M., Witt, L. A. & Harris, K. (2010). P = f (Ability x Conscientiousness): Examining the facets of conscientiousness. *Human Performance, 23,* 343-360.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*, 322-338.

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D. and Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching, 49,* 363–393. doi: 10.1002/tea.21005

Reiner, C., Proffit, D. R. & Salthouse, T. (2005). A psychometric approach to intuitive physics. *Psychonomic Bulletin and Review, 12,* 740–745.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development, 44*(2), 43-58.

Roberts, B. W., Kuncel, N., Shiner, R., N., Caspi, A. & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socio-economic status, and cognitive ability for predicting important life outcomes. *Perspectives in Psychological Science, 2,* 313-345.

Roberts, B.W., Chernyshenko, O. S., Stark, S. & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58,* 103–139.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503-524). Charlotte, NC: Information Age Publishers.

Shute, V. J. & Ventura, M. (2013). Measuring and supporting learning in games: Stealth assessment. Cambridge, MA: The MIT Press.

Sweetser, P. & Wyeth, P. (2005). Gameflow: A model for evaluating player enjoyment in games. *ACM Computers in Entertainment*, *3*(3), 1-24.

Tobias, S. & Fletcher, J. D. (Eds.) (2011). *Computer games and instruction*. Charlotte, NC: Information Age Publishers.

Ventura, M. & Shute, V. J. (2013). The validity of a game-based assessment of persistence. Computers in Human Behavior, 29, 2568-2572.

Ventura, M., Shute, V., Wright, T. & Zhao, W. (2013). An investigation of the validity of the virtual spatial navigation assessment. *Frontiers in Psychology, 4:852.*

Ventura, M., Shute, V. & Zhao, W. (2012). The relationship between video game use and a performance-based measure of persistence. *Computers and Education, 60,* 52-58.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

White, B. Y. (1994). Designing computer games to help physics students understand Newton's laws of motion. *Cognition and Instruction, 1*(1), 69–108.

Wilson, K. A., Bedwell, W., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J., … Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation & Gaming, 40*(2), 217–266.

Metacognition and Self-Regulated Learning

B. Goldberg, Ed.

CHAPTER 9 – Metacognitive Supports to Drive Self-Regulated Learning Experiences

Benjamin Goldberg
U.S. Army Research Laboratory – Human Research and Engineering Directorate

Introduction

To put it simply, recent developments in computer-based technologies are changing the way people learn and develop academically, as well as professionally. Technology is being used in the classroom more than ever, with new tools and methods completely reshaping how people interact with learning content and materials (i.e., interactive e-textbooks distributed to students on Apple iPads [Sloan, 2012]). In turn, where people learn is also rapidly changing. With enhanced mobile networks that support on-the-go internet access and the availability of advanced lightweight portable computers, someone can conceivably learn and train from anywhere in the world. This is leading to a culture based around the self-regulation of learning, especially within academic institutions and professional industries like medicine and the military that value continual on-the-job training for skill development. In this context, self-regulated learning (SRL) refers to self-initiated and self-managed instruction beyond the formal classroom environment (Bjork, Dunlosky & Kornell, 2013).As such, recent research strives to enhance technology-based learning environments through the incorporation of tools and methods that promote SRL by embedding strategies linked to metacognitive awareness and regulation. The following section presents current research among leading experts in the field, with recommendations for instructional management and modeling techniques for enhancing systems to monitor and support SRL development in the absence of live instruction. Each chapter concludes with recommendations for functions and capabilities GIFT needs to support for enabling techniques to model and tutor metacognitive behavior with a goal of improving SRL skills.

Strategies to Support Metacognition and Self-Regulated Learning

We learned from chapters in the previous section the role affect and motivation play in the learning process and how ITSs apply strategies to mitigate and promote states based on their effect on performance outcomes. The four chapters in this section highlight the common themes of using technology-based instructional systems to help students become more independent learners. The authors cover research derived from models and constructs linked to SRL, modeling and monitoring techniques to gauge students' cognitive and metacognitive abilities, defined strategies and tactics for guiding and improving metacognitive processes, and implications for developing authoring tools to facilitate monitoring, modeling, and scaffolding metacognitive processes in an ITS. Collectively, the chapters are oriented toward discussing the pragmatic issues associated with supporting metacognition and SRL in ITSs, and how the application of metacognitive strategies can enhance learning outcomes as they relate to improved learning performance and transfer. As metacognition deals with one's awareness of the knowledge and regulation of cognition, it is important to understand the distinctions between these two parts and how they compliment learning within SRL environments that are open-ended in nature.

The chapter by Goldberg and Spain serves as a review of previous work surrounding the application of instructional strategies in educational technology to support SRL and metacognitive development. The chapter provides a perspective overview of theories and models explaining the SRL processes and how they are used to guide and influence pedagogical practices applied within modern ITS environments. A significant amount of research has been dedicated over the past decade with a goal of determining how computers and artificial intelligence can be applied to improve student learning behaviors and instill SRL

skills that transfer across learning experiences. This includes researching the instructional strategies an ITS can support and the effect they have on intended outcomes. As a result, multiple techniques and strategies have been implemented across numerous examples, with each serving a different function in the SRL process. The authors breakdown the problem space by recognizing common themes (i.e., the metacognitive behavior they aim to support) in strategy implementation and empirical research that investigates how these strategies impact performance outcomes of interest.

The chapter by Biswas, Segedy, and Kinnebrew puts forth a modeling methodology for interpreting a learner's metacognitive skill within open-ended tutoring environments. The technique combines theory-derived metrics for assessing a learner's behavior against expected actions as determined through cognitive and metacognitive task analyses with a data-driven approach that mines student actions to determine sequences and patterns that offer insight into how students actually use the environment to solve designated problems. This approach enables a system to refine model implementations by linking action patterns with designated cognitive and metacognitive objectives determined by the task analysis. Data mining techniques can be used to observe patterns and strategies applied across students to assess how the resources and tools of an ITS are being used, and how those sequence of actions result in an outcome. Through this approach, a system can monitor and assess metacognitive behavior and intervene when sequences and patterns are observed that correspond with errors and low performance. By linking observed actions with designated metacognitive behavior, feedback and scaffolding strategies can be focused on metacognitive tactics that should be considered when conducting problem-solving procedures. These strategies can be represented as scaffold templates that can be adjusted to include domain-relevant information as deemed appropriate by the system author.

The chapter by Roll, Wiese, Long, Aleven, and Koedinger provides a well-rounded overview of the scaffolding techniques applied in today's ITS environments that are aimed at helping students attain better SRL skills and behaviors. To better characterize the problem space, the authors summarize scaffolding techniques across their associated forms, the objectives they aim to attain, and the role they play in the learning process. Basing a review around this framework is important, as pedagogical strategies intended to support SRL and metacognition are anticipated to influence subsequent behaviors in a way that is different from traditional domain-relevant feedback. While multiple forms of SRL scaffolding exist and target varying processes, the objective typically stays the same, to instill effective regulatory behaviors that transfer across domains and learning environments. To attain this goal, the role the various scaffolding forms play must be operationalized. The authors address this by defining distinctions between social components of regulation, and how technology facilitates varying roles as a learner progresses through a topic.

The chapter by Lajoie and Poitras outlines non-adaptive and adaptive instructional strategies used to support self-regulation across educational technology platforms, and uses the domain of medical diagnostics and communication to provide examples of strategy execution. Distinctions are made on the type of strategy and the theoretical constructs of SRL they serve and how cognitive tools are used to support metacognitive processes essential for effective regulatory practices. The authors highlight the constructs associated with this medical domain and how different training platforms are used to instill specific behaviors that can be applied in a real-world context. The social components of learning in the diagnostic reasoning domain are also presented and how technology can be used to facilitate varying processes critical to performance. This is an important perspective as it identifies the role technology plays in supporting skill acquisition within a specialized domain. As complex tasks require the organization and dissemination of information, computer-based training platforms require tools and methods that enable students to metacognitively regulate how that information is handled.

Each chapter provides insight into expanding ITS roles beyond supporting domain-specific information alone. Broadly speaking, this section provides potential ITS developer content on the components

required to execute metacognitive tutoring. This includes looking at various modeling approaches that take into account theoretical foundations associated with a domain, along with methods to monitor environment interactions that link to SRL behaviors. In addition, understanding the forms and roles instructional strategies play is important when authoring pedagogical functions.

Metacognition and Domain-Independency

For GIFT to operate outside a laboratory setting as a domain-independent authoring environment, there are a number of research questions that need exploration. One such question is based around GIFT supporting SRL, and the efficacy of defining persistent metacognitive strategies that can be applied across domain applications. GIFT works with system authors by providing instructional strategy recommendations, which are then translated into tactics as they relate to the training context. These tactics are used during ITS runtime and are selected based on a learner's individual differences. At the current moment, feedback in GIFT is domain dependent and requires explicit content linked to each concept modeled. An example would be GIFT requesting a hint for concept 2.1.2, with the tactic linked to that hint being the specific prompt to display. When it comes to metacognitive feedback, what are the implications to a domain-independent approach? First, modeling techniques need to be developed to monitor an individual's practice of metacognitive strategies that can be expressed in a generalized format. An example would be incorporating a combined modeling approach, as described in Biswas, Segedy, and Kinnebrew's chapter, or by adapting a help-seeking model, as highlighted in Koedinger, Aleven, Roll, and Baker (2009). Researching and establishing models based around commonly available GIFT interactions (e.g., request hint button) can be used to build a representation of how effective students use the interface to solve problems and troubleshoot errors. This approach can aid in detecting learners not practicing good metacognitive behaviors and can be used to trigger feedback interventions to improve their understanding of available strategies. With modeling techniques in place, generic tactics can be identified that are based around effective metacognitive behavior. While tactics can be represented in a domain-independent format, monitoring how a learner adapts their behaviors as a result of the intervention is an open question.

A challenge that must be addressed is establishing an authoring environment and workflow for supporting SRL-derived modeling techniques and linking outputs to prescribed strategies that influence the regulation of metacognitive behaviors. The ultimate goal is to support and influence a learner's approach to problem solving and learning in general. While the research identifies multiple examples of successful strategy implementation, what the literature lacks is guidelines for when best to instantiate them based on the domain being trained and the environment the interaction is taking place in. To enhance GIFT's authoring functionalities, a generalized ontology is required that links specific instructional strategies and techniques with high-level domain relevant content, along with the types of tutoring environments and the services they can afford. By defining these relative dependencies, an ITS developer can embed empirically recommended metacognitive tutoring functions based on characteristics associated with the content being produced.

References

Bjork, R. A., Dunlosky, J. & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444.

Koedinger, K., Aleven, V., Roll, I. & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. *Handbook of metacognition in education*, 897-964.

Sloan, R. H. (2012). Using an e-Textbook and iPad: Results of a Pilot Program. *Journal of Educational Technology Systems, 41*(1), 87-104.

CHAPTER 10 – Creating the Intelligent Novice: Supporting Self-Regulated Learning and Metacognition in Educational Technology

Benjamin Goldberg[1] and Randall Spain[2]
[1] U.S. Army Research Laboratory; [2] U.S. Army Research Institute

Introduction

Educational technology is being used like never before. These platforms enable education and training opportunities in non-traditional settings and support a new culture of "learning by convenience." As a result, educational technology is changing the way people learn and develop academically, as well as professionally. However, the effectiveness of educational technology has shown mixed results when users are left to engage with content on their own. Based on this issue, there has been a push by the research community to study regulatory procedures involved in learning and how these processes apply within computer-based learning environments (CBLEs). The theoretical construct guiding this line of study is Self-Regulated Learning (SRL). In its most basic form, SRL characterizes an individual's efforts to monitor and control their own learning before, during, and after an educational event. For the context of this chapter, SRL refers to self-initiated and self-managed educational events that take place outside of formal classroom settings and in unsupervised computer-based environments (Bjork, Dunlosky & Kornell, 2013). From a learning perspective, SRL is an overarching construct that operates on higher-order cognitive skills such as metacognitive monitoring and awareness. This metacognitive ability allows an individual to monitor and assess their own state of knowledge and engage in regulatory actions to help them handle new problems in unfamiliar domains.

A recognized issue with computer-based instruction is that a majority of individuals lack the metacognitive ability to regulate their own learning and problem solving when engaging in new topics. Research surrounding SRL with educational technology has consistently shown most individuals exhibit poor cognitive behaviors and shallow learning strategies when attempting to study on their own (Azevedo & Aleven, 2013; Koedinger, Aleven, Roll & Baker, 2009; Lajoie, 2008; Winne & Hadwin, 1998). To alleviate this gap, ITSs are being defined as major focal points in regulating student interaction across computer-based learning events. In the traditional use case, ITSs are designed to manage and regulate learning experiences within a specified domain. ITSs do this by monitoring student performance and tailoring the educational experience. Tailoring may involve between-lesson adaptation, such as altering the student's path through the instructional modules or within-lesson strategies such as providing error-sensitive feedback, hints, prompts, worked-examples and other forms of support based on real-time student performance.

While shown to be effective in helping individuals gain new knowledge and learn problem-solving procedures in well-defined problem spaces (VanLehn, 2011), a typical ITS confines its pedagogical approach to domain material alone, with little emphasis on promoting metacognitive and self-regulated learning strategies that can be applied across varying learning contexts. Current research strives to enhance such systems through the incorporation of tools and methods that promote SRL by incorporating strategies linked to metacognitive awareness and regulation (see Azevedo and Aleven [2013] for a comprehensive review of current research in the field). This is based on empirical evidence in the learning sciences community showing the benefit of training metacognitive strategies and their subsequent impact on future learning outcomes (Azevedo & Aleven, 2013; Koedinger et al., 2009; Poitras, Lajoie & Hong, 2012; Roll, Holmes, Day & Bonn, 2012). The overarching goal of metacognitive research within educa-

tional technology is to identify strategies and tactics that instill SRL behaviors across a population of learners when no form of instruction and guidance is made available.

The purpose of this chapter is to summarize recognized best practices identified within the literature for promoting SRL and metacognition and provide recommendations for how to incorporate these strategies within the GIFT architecture. In this literature review, we provide a synopsis of previous research investigating SRL and metacognition as it relates to educational technology, with an emphasis on ITS. We begin with a discussion of the theory and assumptions of SRL. This is followed by a focused review of empirical research on instructional strategies used to promote SRL and metacognitive development in ITS and their effect on learning outcome. As the previous section in this book touched on the affective and motivational dimensions of SRL, this review focuses on cognitive strategies that serve to enhance an individual's metacognition in ITSs. The chapter concludes with a discussion of how these strategies can be incorporated into GIFT's development as a means to promote domain-independent learning.

Theory Behind Self-Regulated Learning

SRL describes the process of taking control of and evaluating one's own learning and behavior (Butler, Cartier, Schnellert, Gagnon & Giammarino, 2011). As a higher-order cognitive function, SRL is guided by metacognitive processes (i.e., the knowledge and regulation of one's own cognition), strategic actions and behaviors (i.e., planning, monitoring, and assessing one's own performance), and motivational components (i.e., goal setting and self-efficacy) (Flavell, Miller & Miller, 1985; Schraw, Crippen & Hartley, 2006). These functions allow self-regulated learners to set goals, monitor their progress toward defined goals, and adapt and regulate their cognition, motivation, and behavior in order to reach the specified goals (Anderman & Corno, 2013; Bransford, Brown & Cocking, 2000). In addition to metacognitive components, SRL also accounts for the various behavioral and affective strategies that students bring to a learning event. The management and control of one's effort during learning is based on motivational and affective factors that influence the level of persistence and energy an individual is willing to put forth when learning is left in their own hands (Pintrich & De Groot, 1990; Zimmerman & Moylan, 2009).

Building upon the definition of SRL reported above, models created to explain self-regulatory procedures used during learning highlight students' proactive efforts to develop knowledge and skill. While there are various models of SRL that differ in the choice of constructs and processes applied within (Azevedo, 2009; Pintrich & De Groot, 1990; Winne & Hadwin, 1998; Zimmerman & Schunk, 2011), Poitras and Lajoie (2013) demonstrate that each model shares several basic assumptions. First, many models acknowledge different areas of regulatory activity and control. For example, Pintrich (2004) defines four areas of control, these being: cognition (i.e., the cognitive and metacognitive activates that individuals engage in to adapt and change their cognition), motivation and affect (i.e., the attitudes, beliefs, and perceptions individuals bring to a learning event), behavior (i.e., the actions a learner may engage in to increase learning success such as exhibiting more effort or help-seeking), and context (i.e., the learning environment or task). A second shared assumption is that learners have the ability to control and regulate certain aspects of their own cognition, motivation, and behavior as well as some features of the environment. This assumption does not mean that the individual will or can always control these aspects, rather it means some monitoring, control, and regulation is possible. Third, most models of SRL assume there is some type of goal or standards in which learners compare their performance against in order to determine whether their strategy is working or determine if some type of change is necessary. As Pintrich (2004) notes, the general example of learning assumes students set goals for their learning, monitor their progress toward these goals, and adapt and regulate their cognition, behaviors, and affect as they experience difficulties in reaching those goals. The final shared assumption is that the components of SRL mediate the relationship between personal and environmental factors to impact learning and performance. That is,

individuals' regulation of their, cognition, motivation, and behavior influences their level of learning and achievement (Pintrich, 2004).

In addition to these assumptions, researchers suggest SRL involves a number of interacting cyclical phases associated with learning and problem solving (see Figure 1). These phases are comprised of actions taken prior to (i.e., forethought phase), during (i.e., performance/volitional control phase) and after (i.e., self-reflection phase) the execution of a lesson, problem, or task (Schunk & Schunk, 2005; Zimmerman, 2008; Zimmerman & Campillo, 2003). Within each phase, learners engage in certain processes to achieve their learning goals. In the forethought phase, self-regulated learners set goals and strategically plan their learning. It is also in this phase that students evaluate their self-efficacy for performing the task and calibrate their task strategies to align with these beliefs. Next, in the performance phase, learners engage in processes to optimize their performance, namely, self-observation, monitoring, and control. Self-observation activities require students to assess whether their learning strategies are effective in meeting their learning goals while they evaluate their emerging understanding of the topic. If there is disparity between their learning goals and their current level of understanding, students may modify their plans, strategies, and effort in an attempt to close this gap. For example, students may rehearse, self-explain, or elaborate on newly learned information to form a better conceptual understanding of the material. They may also seek help from other sources of information. These actions of self-control are intended to realign students' performance with their learning goals. In the final phase, learners reflect on their overall performance. This reflection may motivate students to modify their learning strategy or their goals and revisit the content if their original goals were not met.
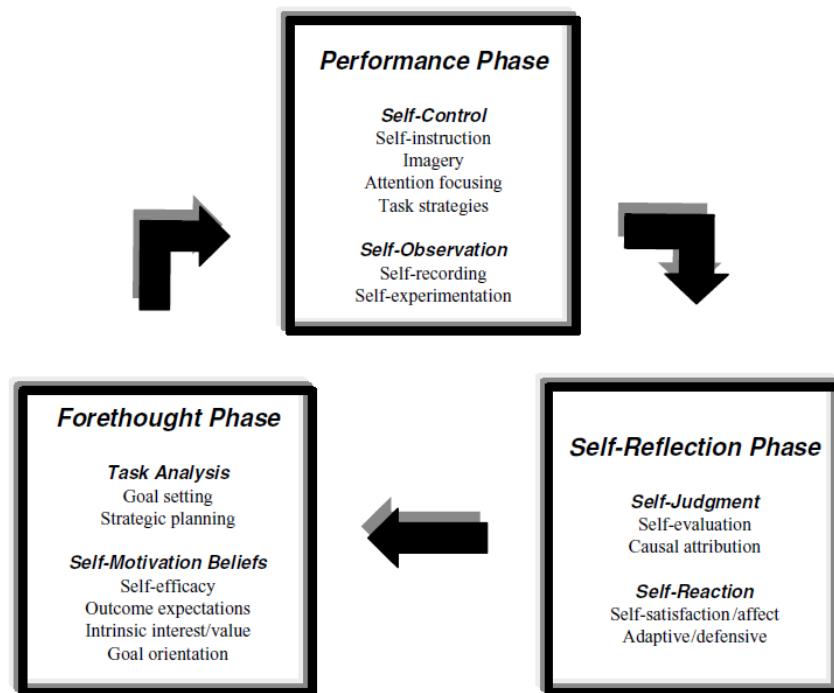
Figure 1. Phases and processes of self-regulated learning as described by Zimmerman & Campillo (2003).

Researchers have used this framework to better understand the various strategies learners use during a learning event. The framework also presents a few basic tenets of what makes the ideal learner. Not only does the ideal learner possess the cognitive abilities required for learning a new skill, but he or she also possesses the metacognitive awareness of what behaviors and strategies to employ, the awareness of what

resources are available and the knowledge of when best to use them, and the motivation to persevere when facing a challenge. This ideal student is commonly referred to as the "Intelligent Novice," in that they use strategies to guide and influence how to approach a problem, and how to adapt their approach based on observable performance outcomes and self-reflection (Bransford et al., 2005).

Theoretical Perspective of Self-Regulated Learning with Intelligent Tutors

While the above section reviews common strategies applied by the ideal learner, the truth of the matter is most learners are far from ideal in the behavioral sense. From an ITS perspective, there is a basic assumption that most users new to a domain need assistance. ITSs are designed to regulate learning experiences by monitoring and guiding system interactions, much as a teacher would in the classroom. The gap motivating this line of research is that current implementations of ITSs often confine their pedagogical approaches to domain-specific feedback and coaching alone, rather than teaching and instilling desired SRL behaviors that can ultimately transfer to other learning contexts.

While elements of ITSs simulate instructor roles, the experience in itself is self-regulated and requires a learner to link system interactions to intended learning goals. This identified co-regulation is very important. While a system regulates its pedagogical interventions based on real-time interaction, a learner will regulate how feedback and coaching is interpreted based on factors linked to relevance and applicability. In terms of co-regulation and metacognitive tutoring with ITS applications, the intent is for a system to guide learners into applying SRL strategies that aid in the understanding of new information and solving novel problems. The goal is for an ITS to facilitate deep learning of a topic while instilling behaviors that support future SRL opportunities. Identifying strategies to promote metacognitive awareness requires an understanding of the processes and phases theoretically linked to SRL approaches. Current ITS research aiming to support metacognitive development base pedagogical techniques on fundamental strategies applied by the ideal learner. These behaviors coincide with the theoretical phases of SRL and serve as moderators for monitoring and regulating one's own learning.

Using Intelligent Tutors to Study and Promote SRL

As previously stated, the problem acknowledged in this review is that people typically don't spontaneously engage in metacognitive and SRL activities. For this reason, there is a push for ITSs to focus on scaffolding SRL and metacognitive behavior to help people become better learners, in general. ITSs provide a unique environment for studying metacognition and SRL behaviors. First, whereas hypermedia environments and online learning are intended to help students learn a complex set of interrelated concepts, ITSs focus on learning by doing and problem-solving tasks. As such, the way in which metacognition is supported in ITSs may differ from the way it is supported in other CBLEs (Azevedo & Aleven, 2013). Researchers have studied how fixed prompts, pumps, and computer-based tools can guide students in evaluating their own learning within ITS (Azevedo & Hadwin, 2005). They have also studied how these forms of support can be applied adaptively so that the type and level of support can be calibrated to the needs of the individual learning. The goal of these metacognitive scaffolds is to help students reach a level of learning that they would not achieve on their own (Pea, 2004; Wood, Bruner & Ross; 1976; Wood & Wood, 1999).

Second, ITSs offer opportunities to unobtrusively collect data that can allow researchers to investigate how learners solve problems, access and use support and help resources in ITSs, and apply metacognitive skills during different learning activities. Third, ITSs allow researchers to vary the design of the environment in order to study the influence of specific instructional strategies on learning outcomes (Azevedo & Aleven, 2013). This last feature is especially important in empirically examining the impact of different instructional interventions on learning outcomes. Finally, ITSs have the ability to model and support

specific SRL and metacognitive activities. That is, in addition to offering support at the domain level, ITSs can also help students become better learners in general by offering feedback and advice on how to use help resources more effectively.

The goal for the remainder of this chapter is to review several instructional strategies empirically evaluated within ITSs, with the discussion being confined to strategies that aid in the cognitive processes of SRL. We use the model of SRL depicted above to guide our discussion and highlight how support in ITSs can be offered in the forethought, performance, and evaluative phases of learning. Rather than provide a high level overview encompassing a majority of the research in the field, we provide descriptions of strategies believed to have the greatest impact on learning based outcomes. In the context of tutor development within the GIFT architecture, the goal is to establish a set of strategies the framework can support and conceptualize their application for determining future development efforts to support their implementation, both from a modeling and pedagogical delivery standpoint. In the following section, we describe several ways in which metacognitive support has been applied in ITSs and how those strategies can be supported with GIFT. For example, with regard to forethought and planning, we highlight how ITSs have been used to promote metacognition though goal setting and strategic planning actions. With regard to the performance phase, we review how different metacognitive prompts and modeling techniques have been used to make students become better regulators of their learning. Finally, with regard to reflection, we review how ITSs have been used to promote student self-reflection through learning by teaching practices. Following this review, we describe design recommendations for GIFT.

Forethought

As noted in Figure 1, a self-regulated learner begins their interaction in the forethought phase by defining and organizing goals associated with the problem space. These goals are defined on a cognitive and motivational level. It is also in this phase that the learner activates prior knowledge associated with the topic and formulates an initial strategic plan aimed at attaining set out objectives. Once a topic is introduced, these processes are linked to the forethought phase of SRL theory-based models and dictate the initial iteration of the performance and self-reflection cycle.

Goal Setting and Strategic Planning

In terms of research supporting metacognitive development with educational technology in the forethought phase, the processes found to be most prevalent in the literature include goal setting and strategic planning. While it is agreed that SRL is a cyclical process and temporal by nature (Bjork et al., 2013), the types of metacognitive processes associated with forethought can be applied across varying levels of granularity. In the typical ITS setting where a learner focuses on solving a set of problems or scenarios, goal setting is critical because it can assist in establishing an initial plan of execution for achieving identified goal states. In addition, established goals serve as criteria for assessing performance. Based on outcomes from an executed plan, a learner will make iterative strategic modifications based on performance-related information, self-reflection, and the recycling of goals in working memory (Azevedo, Cromley & Seibert, 2004). For highly effective students, goals are organized in a hierarchical fashion (Bandura, 1991; Zimmerman & Campillo, 2003). Once a task is defined in an operational context, an effective learner decomposes the problem space into target goals and sub-goals that explicitly identify what is required to reach the defined objective (Pintrich, 2002; Winne, 2001; Zimmerman, 2002). This task deconstruction enables the learner to use identified goals as moderators for retrieving and activating prior knowledge in memory (Azevedo & Cromley, 2004). Goal setting can also be used to identify gaps in knowledge that require further attention and the execution of help-seeking behaviors.

Regardless of the task, goal setting assists a learner in defining, on a cognitive level, what exactly needs to be done to complete a problem. This representation is used to construct an initial strategic plan for execution within the performance phase of SRL. The plan is composed of problem-solving strategies a learner is metacognitively aware of (i.e., using an analogy to compare a current problem against a similar experience stored in memory) and knowledge of the environmental resources available for assistance (i.e., using a calculator to perform a complex mathematical expression or a glossary to lookup a theorem). When a strategic plan is put into action, an outcome is produced that dictates how to appropriately proceed. It is during this self-reflection phase that an effective learner is able to link actions executed in the environment with the performance outcome information made available. When outcomes do not meet defined target goals, the original strategic plan is modified based on the extent of the error and the learner's ability to identify root causes. This might incorporate an immediate reattempt based on a recognized mistake or exploring available resources to obtain new information linked to the problem.

While the benefits of goal setting and strategic planning are made clear in the SRL literature, in reality, the majority of novice learners lack the knowledge and skill to properly perform these strategies under novel contexts incorporating educational technology (Azevedo, 2009; Hadwin & Winne, 2001; Lajoie, 2008). Rather than determine why people are bad at this, the real question is how can ITSs be used to improve a learner's metacognitive skills associated with forethought processes? From the ITS perspective, these metacognitive strategies are generally offline practices, in that they do not produce data inputs for run-time performance models, making it difficult to assess capability and provide focused instruction on their application. In terms of previous research examining the use of instructional strategies to enhance goal setting and strategic planning processes, studies from the past 10 years focused on either the utility of upfront instructional materials for just-in-time metacognitive training or the use of learning-theory based interaction models that require a learner to apply specific forethought strategies that dictate how a problem will be solved.

For just-in-time training with upfront materials, one camp of researchers explored various approaches for instilling SRL processes prior to a learning event with educational technology. Azevedo et al. (2004) performed multiple empirical studies examining the effect differing types of metacognitive training materials have on performance outcomes when individuals learn from a self-regulated hypermedia environment. An initial study conducted by Azevedo and Cromley (2004) compared performance marks and shifts in mental models across two conditions of students learning about the circulatory system. The experimental group received an upfront 30-minute training session on the use of SRL strategies that were empirically based and found to help foster deep conceptual understanding of a topic, while the control condition received no SRL-focused training. Results showed the SRL training condition led to significant increases in understanding of the circulatory system when compared against the control. Evidence was found supporting those receiving SRL training to effectively use SRL strategies and processes that led to larger shifts in mental model assessment outcomes.

The caveat with this study is that the 30-minute training session was administered by a human proctor. While this contradicts the chapter's goal of identifying strategies managed directly by an ITS, this research provides valuable insight into mechanisms that can easily be applied prior to any learning session. Despite the material being delivered by a human instructor, the content did not vary between students. Each proctor was provided a script listing the numerous SRL-based strategies along with instructions for how to apply them. While the learner benefited from the direct interaction with a human, the same material could conceivably be provided to the learner as a reference resource embedded within the system. The researchers concluded by recommending the use of a standardized script for upfront SRL training and to extend a system to include metacognitive scaffolding for use during performance (e.g., activate prior knowledge through prompts) (VanLehn, Siler, Murray, Yamauchi & Baggett, 2003).

In a complimentary study examining the effect of scaffolding techniques within a hypermedia learning environment, Azevedo et al. (2004) demonstrated the ineffectiveness of providing upfront materials that assist a student in defining a hierarchy of goals associated with content based on leading questions. The study involved three conditions: (1) adaptive scaffolding provided by a human tutor who assisted a student in applying metacognitive strategies, (2) a fixed scaffold approach that included only a reference printout of goals and subgoals associated with the domain being learned, and (3) no scaffolding. Results show the individuals receiving the adaptive scaffolding demonstrated a deeper conceptual understanding of the circulatory system based on a pre-/post-test comparison analysis. While it was hypothesized that presenting learners with an explicitly defined list of goals would help guide their behaviors in the learning environment, the outcomes show those receiving the fixed scaffolds performed in relative comparison to the control group. While it is recognized that goal-setting is an important component of SRL, research has shown little progress in tools and methods applied in educational technology to assist novices in effectively performing these strategies. While setting goals requires the ability to deconstruct a problem into target objectives, understanding how those goals can be applied to guide learning behaviors is less understood, as made evident by this second study. Further research is needed to examine approaches for training goal-setting strategies and how those goals dictate strategic planning.

Problem-Solving through Back-Chaining with the Target Variable Strategy

Another approach linked to processes conducted in the forethought phase of the SRL cycle was studied by Chi and VanLehn (2010). They developed a statistical probability ITS called Pyrenees that requires a student to follow a generic problem-solving procedure that incorporates metacognitive behaviors when formulating a solution. The procedure is based on a technique called the target variable strategy (TVS), a domain-independent back-chaining procedure that systematically breaks a problem down into its constituent piece parts. TVS is composed of three phases: (1) translate a problem statement into end-state objectives, (2) apply principles and generate equations for reaching the defined objectives, and (3) solve the selected equations. The first two phases of this approach are associated with goal-setting and strategic planning, while the third is associated with execution of the devised plan. Based on how the problem is presented, a student is required to conceptually define what variables are known and what variables are sought after. From there, a student identifies one of the sought variables as the target variable to build a strategic plan around. This target variable is intended to activate prior knowledge that can be used to create a solution. A solution is generated by defining equations that contain the target variable for solving. To test the effectiveness of the TVS strategy in promoting efficient problem-solving behaviors, Chi and VanLehn (2010) ran a study looking at students learning the same topic from two separate ITSs. The experimental condition involved Pyrenees, which taught probability and requires a learner to solve problems by engaging in the TVS procedure, while the control group studied probability with the Andes ITS, a tutoring system that did not teach or require a specific strategy (VanLehn et al., 2005). To examine if students applied the TVS strategy in a transfer setting following instruction, participants interacted with the Andes physics ITS as a second transfer domain.

Prior to analysis, the researchers classified individuals as high-capability or low-capability students based on their math SAT score. This is used to determine if training the TVS strategy has more of an effect based on a student's pre-established ability within a domain. Results from this study show the low-capability learners in the experimental condition scored higher on all assessments when compared against the low-capability control condition, while both experimental groups outperformed the high-capability students in the control (Chi & VanLehn, 2010). Both test scores and dynamic assessments show that low-capability students catch up with the high ones when interacting with the experimental condition but not in the control. On some measures, the low experimental learners surpassed the high control ones. Data suggest a main effect of teaching low students the TVS produced a better focus on domain principles during learning, while high ability students, regardless of the condition, showed application of TVS

procedures. Overall, the manipulation shows an aptitude-treatment interaction. The metacognitive skill of focusing on individual principle applications is found to be a beneficial strategy to teach low learners.

Improving Self-Assessment Skills

Another approach researchers have examined in an attempt to improve students strategy selection is supporting students' self-assessment skills. Self-assessment refers to the tendency and ability to accurately evaluate one's own knowledge while learning. Research shows students often overestimate their ability when it comes to assessing their own level of knowledge; many students base their ability estimates on their familiarity with the topic. These erroneous assessments can lead to suboptimal strategy selection. Accurate self-assessment, on the other hand, has been shown to correlate with productive help-seeking behaviors. Currently, only a small number of systems provide support for self-assessment in order to help students choose the appropriate cognitive strategy and monitor their progress (Bull & Kay, 2007; Gama, 2004). The goal of such an intervention is to help students become more aware of their relative strengths and weakness of their knowledge in relation to a learning task. There is much less evidence of explicitly supporting self-assessment in ITSs. To meet this need, Roll, Aleven, McLaren, and Koedinger (2011) developed a self-assessment tutor with the goal of helping students improve the accuracy of their self-assessments, and use their self-assessments to inform strategy use. The tutor asks students several questions regarding their current level of knowledge for the topic, in which students respond using a menu-based interface. For example, the tutor begins the self-assessment process by asking students to predict whether they could solve a given target problem without making errors. Students are then asked to solve the target problem. After solving the problem, students are asked to recall their initial self-assessment and reflect on whether they solved the target problem without making any errors. Then they are asked if they correctly evaluated their knowledge and whether they would ask for support next time. Though rudimentary in its approach, research has shown students improve their ability to identify their strengths while working with the self-assessment tutor. In addition, research shows students transferred the improved self-assessment skills to unsupported problems in the same tutoring environment (Roll et al., 2011). These research results are promising and show that relatively simple interventions in an ITS interface can help students become more aware of their strengths with regard to problem solving. Helping students become more aware of their abilities is especially important as this initial appraisal strategy selection and subsequent performance.

In summary, this section focused on research examining multiple approaches for training forethought processes that are believed to influence problem-solving behaviors. The strategies reviewed include just-in-time upfront SRL training, goal-setting aids, back-chaining through the TVS, and self-assessment prompts. Each strategy, with the exception of goal-setting aids, was found to impact learning outcomes when controlled in an experimental setting. This conveys the influence forethought processes have on the subsequent performance phase, and provides support for further research into identifying approaches for incorporating these mechanisms in GIFT. The next section discusses research surrounding metacognitive strategies enacted in the performance phase of the SRL model that involves monitoring performance and regulating interaction based on observed outcomes.

Monitoring and Regulating Performance

In the performance and reflection phases, metacognitive strategies are selected based on characteristics of the learning task and how outcomes compare with defined goals. For example, if a student is reading a passage with the goal of retaining in memory the fundamental concepts and theories linked to the topic so as to recall them when solving a problem, then the student may engage in regulatory actions that force them to express the received information in their own words. Self-explaining content in one's own

thought allows an individual to strengthen relationships that are understood on a knowledge component level and helps identify concepts that require further assistance in comprehending.

If the learning task requires problem solving, which is most often associated with ITS applications, then there are a number of different metacognitive decisions that a student must make when solving the problem. For starters, when learners approach a problem, they must determine whether they have the ability to solve the problem. If they are familiar with the problem, they may try to solve it on their own. If they are unfamiliar with the problem, they may consult their book or seek help from another resource. Many ITSs offer built-in support, such as online glossaries, hints, and feedback. If a student solves a problem and the resulting outcome is incorrect, then the individual may use an ITS's hints or feedback to guide error detection and trace the source of the error. In the instance where a learner produces a correct outcome, an ideal student will perform self-explanation practices in the reflection phase before moving onto the next problem. A learner will attempt to explain how he or she reached the solution to ensure the learner has complete understanding of how the answer was calculated. If an impasse is reached during this reflection phase, goals are subsequently adjusted so as to identify a remediation approach. In this case, a student may choose to solve more problems or carefully review the concepts underlying the problem.

Regardless of the type of impasse identified, either through self-explanation, self-correction, or failure to find a solution, an effective student will perform help-seeking behaviors that use available resources to obtain information that may help create a better understanding of the concept being focused on. For these types of metacognitive strategies, a learner must know the resources available for guidance, where they are located and what resources are most appropriate to use based on the context of the issues being faced. In the following sub-sections, we briefly review how different types of SRL and metacognitive scaffolding, mainly self-explaining, error-detection and self-correcting, and help-seeking have been effectively modeled and supported in ITSs.

Fostering Metacognition through Self-Explanations

One of the most well-researched metacognitive strategies for improving student learning is prompting students to self-explain as they solve problems or study lessons. The benefits of self-explaining while studying or solving problems have been investigated by a number of researchers, starting with the work of Chi and VanLehn (2010) who found successful learners tended to explain or engage in more generative processes (elaboration, paraphrasing, etc.) than unsuccessful learners – a phenomenon they called the self-explanation effect.

Research on the self-explanation effect in ITSs has shown strong benefits in which students who are prompted to engage in self-explanation perform better on subsequent tests than students who do not self-explain, particularly in well-defined domains such as mathematics, physics, and biology.

For example, Aleven and Koedinger (2002) incorporated self-explanation prompts into a step-based ITS designed to help students solve geometry problems. The ITS required students to select from an onscreen menu the explanation that best justified their answer as they worked through problems. After submitting their answer and explanation, the ITS provided students with feedback on the correctness of their inputs. If students solved the problem correctly, they advanced to the next problem. If they erred, they were required to input the correct answer before advancing. In two studies, Aleven and Koedinger (2002) found students who were prompted to self-explain their problem-solving steps acquired a deeper understanding of the geometric principles and theorems compared to students in the control condition who were not prompted to self-explain. In particular, students in the self-explanation condition were better able to justify their answers and solve unfamiliar, but related, problems than students who did not self-explain.

Aleven and Koedinger (2002) concluded that students who were channeled to self-explain acquired a deeper and more meaningful understanding of the geometric principles and theorems, and acquired less shallow procedural knowledge surrounding these concepts than students who did not self-explain.

Atkinson, Renkl, and Merrill (2003) found similar benefits of prompting students to self-explain as they solved problems in a CBLE. Similar to Aleven and Koedinger (2002), Atkinson et al.'s (2003) CBLE required students to select from an onscreen menu the principle that best justified their solution as they solved probability problems. Students also received feedback on the correctness of their solutions and explanations. Results showed students who were prompted to self-explain not only performed better on post-test items that required an understanding of simple rules for solving probability problems, but they also performed better on novel items that required a deeper understanding of these rules compared to students who did not self-explain.

In addition to having students self-explain by selecting their reasoning from an onscreen menu, researchers have investigated the benefits of having students compose their own explanations. One might assume that requiring students to type their own explanations, rather than choose from a reference list, would lead to better learning as this technique requires students to generate their own explanations. However, research on interfaces that support free-text entry suggests feedback on self-explanations may be an important factor in their success (Koedinger, Aleven, Roll & Baker, 2009), which requires accurate assessment of free-text entries as they relate to a desired input.

For instance, Aleven and colleagues (2004) compared the effectiveness of two versions of an ITS that supported free-text entry of self-explanations; one that did not provide feedback (Aleven & Koedinger, 2000) and one that provided feedback through natural language dialogue (Aleven, Ogan, Popescu, Torrey & Koedinger, 2004), both of which were compared to an ITS that used menu-driven prompts. Aleven and Koedinger (2000) found in the absence of feedback students frequently ignored prompts and provided very few good explanations. However, when the tutor provided feedback, students self-explained and the quality of their explanations improved considerably (Aleven et al., 2004). In fact, Aleven et al., found students who received feedback on their text-based self-explanations performed just as well on measures of retention and transfer as students who self-explained using a menu-based prompt. More importantly, these researchers found students became better at stating their explanations as they progressed through the lesson, which led the researchers to conclude there were some benefits of eliciting self-explanations using the natural language interface. There was also causal evidence suggesting that better feedback from the tutor led to greater progress in making accurate explanations. These findings suggest that providing feedback may be a key design consideration when prompting students to self-explain using a natural language dialogue interface.

Further evidence for the importance of providing feedback in free-from explanations may be drawn from the work of Johnson and Mayer (2010). These researchers examined the benefits of self-explanation in a game-like environment. Students were required to answer questions about electrical circuitry and explain the reasoning for their answers by either selecting the appropriate principle from an onscreen menu or typing the principle into an onscreen box. Participants earned points for correctly solving the circuit problem, but did not receive any feedback on the correctness of their explanation. Results showed significant benefits of self-explaining when students used the menu-driven, but not when they typed their self-explanation using the onscreen textbox. In fact, there was no difference in performance between students who typed their explanations and those in the control group who did not self-explain at all. Johnson and Mayer (2010) concluded,

> "Adding self-explanation using a selection format was successful because it fostered essential and generative processing (by requiring the player to think about explanations for what was happening in the game), while minimizing extraneous processing (by maintain-

ing some semblance of game flow). In contrast, adding self-explanation using a generative format was not successful because it reduced the motivating features of the game by greatly disrupting game flow and created extraneous cognitive processing aimed at composing and typing text (pg. 1251)."

Another factor that may have contributed to the ineffectiveness of the text-based self-explanations was the absence of feedback. In this study, students did not receive feedback on their explanations; they only received feedback on their answers. However, Aleven et al., (2004) showed providing feedback on generated self-explanations was critical to their success. Future research should continue to examine the moderating role of feedback when coupled with different self-explanation techniques.

The studies mentioned above all applied self-explanation prompts in a fixed manner; that is each student received the same type of prompt at the same time. Fewer researchers have attempted to incorporate adaptive self-explanation prompts into CBLEs. One notable example is the work of Conati and VanLehn (2000) who developed an ITS that prompted students to self-explain based on estimates of student ability and performance. These researchers found the adaptive support was successful in helping students acquire domain knowledge, especially in the early stages of knowledge acquisition. Another example comes from the work of Weerasinghe, Mitrovic, and Martin (2009) who designed a general model for supporting individualized self-explanation by engaging students in tutorial dialogue. Students were adaptively prompted to explain their problem-solving steps based on errors they made. Results showed that self-explanation prompts based on students domain knowledge improved domain-level learning. Conati (2009) describes another means for adaptively prompting students to self-explain. Specifically she describes two different ITSs that largely try to determine when to provide self-explanation prompts, and which type of self-explanation prompt to provide, by relying on a probabilistic student model that examines how long students spend studying relevant examples and how well they know the domain principles. The model also considers how likely students are to self-explain when determining which intervention to apply. In summary, researchers are exploring how to adaptively apply self-explanation prompts in ITSs; however, the evidence is limited as to which techniques are most successful.

Enhancing Metacognition with Cognitive Tools

A different perspective on ITS design is seen in BioWorld, an interactive CBLE that trains medical practitioners on diagnostic reasoning across an array of simulated exercises (Lajoie, 2009). BioWorld was developed to support expertise development of medical diagnosis by assisting students with externalizing and evaluating their reasoning processes. The system is designed around social cognitive theory and models of cognitive apprenticeship, where the model accounts for dependencies between the student, the teacher, the context of instruction (i.e., the materials and resources available), and the associated assessments (Brown, Collins & Newman, 1989). In the context of SRL and complex problem solving with educational technology, models of cognitive apprenticeship are used in the design phase to identify the aspects of instruction a system must support and the types of tools required to do so. This is based on understanding the processes for solving a problem (e.g., domain modeling) and providing scaffolds to aid in obtaining a solution, including aids to assist with metacognitive processes. ITSs designed around the tenets of cognitive apprenticeship focus user interaction around a set of cognitive tools that manage the delivery of domain materials and practice opportunities through modeled representations of student knowledge and comprehension across a problem space, along with scaffolds that guide interaction based on models of human behavior (Lajoie, 2008).

The most explicit human behavior linked with scaffolding and cognitive apprenticeship is the externalization of expert knowledge through verbal protocol methods (Lajoie, 2009). This strategy defines on a tacit level the cognitive and metacognitive processes an expert applies when solving a problem, and ultimately,

is used as a basis for a learner to observe and enact in subsequent practice opportunities (Lajoie, 2009). For educational technology, cognitive tools are used to replicate the externalization of knowledge by providing tools and processes that are inherently used by an expert when solving a problem. In the domain of diagnostic reasoning, effective learners often follow a generic self-regulated process of developing hypotheses in forethought, developing a plan to test generated hypotheses, and reflecting on produced outcomes once a plan is put in action. In the instance of BioWorld, the platform tasks users with diagnosing virtual patients based on evidence collected from a patient summary (i.e., the problem description) and by running supported clinical tests (McCurdy, Naismith & Lajoie, 2010). BioWorld provides a mature operational environment where users have the ability to read patient cases, record initial hypotheses for testing, record supporting evidence, access a library for information relevant to the case, and ask for assistance from a help agent. A typical case involves the following stages: the investigation phase where a user identifies relevant evidence and run clinical tests that are potentially useful, a prioritization phase where a user selects the relevant information that support their final diagnosis, and a summary phase where a user produces a free-text narrative describing the important information linked to the case.

While the tasks in BioWorld are well defined by nature, determining a diagnosis based on discriminate information is difficult. Diagnostic reasoning requires metacognitive processes to regulate what information is processed and how that information is processed, and monitor progress toward a solution based on collected evidence. To this effect, BioWorld offers a set of cognitive tools that act as scaffolds for guiding a student to reach an eventual diagnosis. These tools are designed to support monitoring processes along with providing help-seeking resources commonly used during medical diagnostic events (Lajoie, 2009). One such tool embedded in the environment is termed the "evidence palette" as it provides a notebook interface to record information deemed to be important for supporting a diagnosis.

As these tools are provided in BioWorld to support learning processes, there is no guarantee a student will use them, let alone use them correctly. Based on this, McCurdy et al. (2010) ran a study to examine the use of the evidence palette as a metacognitive tool and observe if it is being used in its intended way. Data were collected across two groups of individuals, medical students and expert physicians. This enables analysis to determine not only if the tool is being used, but whether it is being used differently based on the ability and expertise of the individual interacting with the system. Each participant completed three cases, while metrics were collected that corresponded with how the palette was used across the phases of investigation, prioritization, and summary. Results showed students and experts used the evidence palette in distinctly different ways, where experts collected significantly more evidence during the investigation phase (McCurdy et al., 2010). The authors associate this with Glaser, Chi & Farr's (1988) assertion that experts spend more time qualitatively analyzing a problem in the initial stages, while novices attempt to solve a problem immediately.

Supporting Error Detection and Self-Correction

Another important metacognitive skill that researchers have begun to support in ITSs is error detection and self-correction. Early work in this area stemmed from recasting the debate regarding feedback timing in ITSs (Mathan & Koedinger, 2005). As previously mentioned, ITSs usually provide students with feedback as soon as they diverge from the model of expert performance. Though researchers have shown the benefits of immediate feedback, particularly with regard to improving learning efficiency in ITSs (Corbett & Anderson, 2001), some have been critical about its use in the context of CBLEs. Specifically, researchers have cautioned that immediate feedback detracts from students developing important metacognitive skills needed for problem-solving tasks. They also argue that students may use immediate feedback as a crutch when solving problems as opposed to using evaluative skills to understand why they erred (Bjork, 1994; Nathan, 1998; Weerasinghe, Mitrovic & Martin, 2007). However, rather than framing the question as one about feedback timing, Mathan and Koedinger (2005) suggest the more important

question is to determine the "model of desired performance" (Mathan & Koedinger, 2005). If the objective is for students to work through the problems as efficiently as possible, then providing students with immediate feedback might be the best course of action (Corbett & Anderson, 2001). However, if the objective is for students to become better self-regulated learners, then developing a system that supports students in identifying and correcting their errors might be more helpful than providing immediate assistance.

To test this notion, Mathan and Koedginer (2005) developed two version of an ITS designed to teach students how to perform different calculations in a spreadsheet program. One version was based on a model of expert performance and provided students with immediate feedback as soon as they erred. The second system was based on an intelligent novice model and allowed students to make a few mistakes before intervening. This alternate version stepped students through a process of error identification and error correction as opposed to providing immediate corrective feedback. The researchers noted that by helping students reflect on their performance, the outcomes of their performance and how those outcomes were different from the goals, the tutor helped students exercise important metacognitive skills. Results of their experiment showed students in the intelligent novice condition performed better on both immediate and delayed transfer tasks compared to students in the expert model condition. Learning curve analyses also indicated that students in the intelligent novice condition learned at a faster rate compared to students in the expert model condition. The results of this study show that explicitly modeling metacognitive skills and using these models to scaffold student performance is likely to lead to more effective and efficient learning than approaches that merely give students an opportunity to apply these skills (Mathan & Koedinger, 2005).

Promoting Help-Seeking Behaviors

Help-seeking is considered to be an important self-regulatory skill that good learners use to master learning material (Pintrich, 2004). Help-seeking is defined as the ability solicit help when needed from a teacher, peer, textbook, manual, online help system, or Internet. Research on learning in social settings, such as classrooms, shows that help-seeking is an important behavior for independently mastering skills (Karabenick & Newman, 2006). Good learners know when to seek help, what types of resources to seek for help, and how to appropriately rely these resources to overcome any impasse they may reach when learning new material. Unfortunately, research has also shown that those who need help the most are the least likely to ask for it.

Recently researchers have begun to study help-seeking behaviors in the context of interactive learning environments such as ITSs and CBLEs. As mentioned previously, and discussed more thoroughly in several chapters in this book (see Chapters 9, 10 and 13), CBLEs and ITSs offer different forms of support to students. They may offer on-demand hints that give students advice on what to do next at any point during their problem-solving activity or online glossaries that students can use to look up relevant information. Given that many of these learning environments offer some form of online help or support, it's reasonable to assume that appropriate use of these resources would improve learning outcomes. However, recent research indicates that students often misuse help resources. They either ask for support when they do not need it (help abuse) or fail to ask for help when they do, such as after making repeated errors on a problem-solving step (help disuse). These behaviors have been shown to be associated with poorer learning in CBLEs (Aleven, Mclaren, Roll & Koedinger, 2006).

In an attempt to help students become better help-seekers, researchers have begun developing models that allow systems to tutor students on how to rely on help facilities more appropriately. In particular, Aleven et al. (2006) have developed a normative model of help-seeking behavior that can be integrated into cognitive tutors to teach students how to use help resources. This effort differs from other efforts de-

scribed in this chapter in that the approach does not involve prompting students into engage in one specific type of metacognitive behavior (e.g., self-explanation), rather the goal is to have students internalize appropriate help-seeking behaviors that can be generalized from domain to domain. Also, because the tutor does not focus on teaching domain-level knowledge, but on help-seeking behaviors, the tutor can be applied to other domains without much adaptation.

The help-seeking model provides a representation of the preferred metacognitive behavior when working with a step-based cognitive tutor. It specifies that students should work deliberately to solve problems and should seek help only when they reach an impasse in their understanding. Specifically, the model states students should only use help when they don't know how to solve a problem; they do not have a clear sense of what to do next, and they have made an error they do not know how to fix. The model also notes the choice of which help resources students should use, the tutor's on-demand hints or online glossary, is dependent upon students' self-assessed level of understanding. The less familiar the step, the more contextualized the help should be. A student who is unfamiliar with a problem should first ask for a hint, as this would provide the student more contextualized information for solving the problem. A student, who is somewhat familiar with the problem-solving step, but still not entirely sure, should seek help from the online glossary. The model also takes into account how much help a student should solicit, noting that a student should only ask for as many hint levels as needed to get a clear sense of how to solve the problem.

In addition to the model, the researchers created a help-seeking agent that uses the help-seeking model to provide students with contextualized metacognitive feedback as they solve problems. If a student violates the rules of the help-seeking model by engage in metacognitively inappropriate behavior, then the agent gives them contextualized feedback messages in response to such behaviors. The agent uses estimates of student mastery to determine what type of help-seeking behaviors are appropriate for a given student. The model also includes an extensive database of inappropriate metacognitive behaviors, such as clicking through hint levels too quickly, which allows the tutor to provide tailored feedback messages to students.

One of the goals of the help-seeking model is to improve students' help-seeking behaviors in new domains while at the same time, helping students acquire domain-relevant knowledge (Roll, Aleven, McLaren & Koedinger, 2011). Thus far, research has provided mixed evidence of this goal. Roll et al. (2011) integrated the help-seeking tutor into a commercial tutoring system for geometry and found that the help tutor improved students help-seeking behavior while learning geometry. There was also some evidence to suggest that students transferred their improved help-seeking skills when learning content in a new domain. Specifically, Roll et al. (2011) found students who had previously received feedback on their help-seeking behavior spent more time focusing on the tutor's hints and asked for fewer hint levels compared students who did not interact with the help-seeking tutor. Unfortunately, the experimental design precluded the authors from determining whether the effects were the direct result of receiving feedback from the help-seeking tutor or other extraneous variables that were introduced in the study.

Researchers are continuing to explore how ITSs can be used to support effective metacognitive and self-regulatory behaviors during problem solving. The studies described above note several effective techniques. However, further research is needed to determine if the effectiveness of these approaches may be moderated by the learning task or the presence of other system and environmental features. Taken together, these studies shed light on the types of support that need to be incorporated into GIFT's tutoring interface so that researchers can further explore the effectiveness of different metacognitive prompts and scaffolds.

Next, we describe how ITSs have been designed to support techniques that enact learners as the instructor of a topic. Many of the strategies described above involved some level of self-reflection and could therefore be considered adequate approaches for this phase of SRL. In this next sub-section, however, we

focus on a particular learning strategy that has been used and modeled in ITS called learning by teaching. Learning by teaching is a co-regulated learning strategy in which the student becomes the teacher. In this new role, students are required to monitor their students' level of knowledge and reflect upon their own and their students' knowledge.

Learning by Teaching

Learning by teaching with ITSs puts a unique twist on the common learner-ITS interaction. In this paradigm, students teach artificially intelligent agents about a topic. Through this interaction, students are theorized to learn about the topic themselves (Biswas, Leelawong, Schwartz & Vye, 2005). This approach is backed by the cognitive sciences in that teaching others changes how someone approaches a task and can be a powerful tool in priming an individual to develop deep conceptual understanding of a domain. This is supported by research conducted by Biswas et al. (2004) where they reported individuals prepping to teach a topic put extra pressure on themselves to organize the material effectively. To teach, one must gain a deep conceptual understanding of materials and then structure that knowledge in a form they are comfortable in sharing with others (Bargh & Schul, 1980). This process is self-directed and open-ended in nature, where a learner is left to explore, integrate, and structure knowledge before they can assist another in doing the same (Wagster, Tan, Wu, Biswas & Schwartz, 2007). Beyond preparing to instruct, the act of teaching in itself also requires active self-direction and encompasses three critical aspects of learning interactions: structuring, taking responsibility, and reflecting (Biswas et al., 2005). While there is a plethora of research in the field on teachable agents and computer-based learning environments, for this review, we focus on recent work surrounding the ITS Betty's Brain and the various studies examining instructional strategy techniques and their effect on performance outcomes.

Betty's Brain is an ITS that uses teachable agents and was developed to teach middle school students about topics related to earth science (Leelawong & Biswas, 2008). Students teach Betty about a particular topic (such as the food chain, photosynthesis, or the waste cycle) by linking concepts to one another on a simplified visual representation called a concept map. Betty is then assessed following this interaction to determine how much she has learned, where her answers are produced through qualitative reasoning methods based on the chains of links constructed in the student's concept map (Biswas et al., 2005). Following this assessment event, a student is forced to engage in self-monitoring and reflection activities to gauge what Betty knows and identify the concepts where she could use some help. The test results are used by a student to refine Betty's "brain" by modifying the established concept map to represent revised relationships based on the information provided from the assessment outcomes. This is followed by a re-examination to assess if an accurate understanding of the topic is represented. This cycle of testing and remediation continues until Betty successfully passes the administered test. From initial examinations, research has demonstrated students who tutor Betty gain a deeper understanding of material when compared against individuals learning the same content from a traditional ITS instantiation (Biswas, Schwartz & Bransford, 2001).

To further enhance the effectiveness of Betty's Brain, a study was conducted looking at the effect of added metacognitive supports intended to help students demonstrate and develop effective learning behaviors (Wagster et al., 2007). This was accomplished through the development of an SRL-based Betty's Brain that incorporated metacognitive scaffolds to aid students in developing and applying monitoring and self regulation strategies (Tan, Biswas & Schwartz, 2006). It was hypothesized that a teachable agent environment combined with scaffolding and feedback functions would provide opportunities for students to develop metacognitive knowledge and control, resulting in improved subsequent learning (Wagster et al., 2007). These scaffolding and feedback functions were encoded in Betty's persona, where the agent would spontaneously convey metacognitive knowledge at times to aid a learner in developing and applying self-regulation and monitoring strategies.

With Betty's new capabilities, a study was designed to determine if the addition of metacognitively based strategies would enhance the effectiveness of the teachable agent system. The resulting experiment involved three conditions: (1) students taught by an agent, (2) students taught the agent directly, and (3) students taught the agent directly and received metacognitive support while doing so. Data were collected over the course of two weeks as participants worked to create concept maps on aquatic ecosystems. Eight weeks following the main data collection, a preparation for future learning (PFL) study was conducted to determine if the use the learning strategies transferred following interaction with Betty's Brain (Tan, Wagster, Wu & Biswas, 2007). The metrics used to determine the effectiveness of the conditions included data linked to performance outcomes, such as the concept map quality measure, along with the coding of student behaviors and sequences to determine how strategies were applied during interaction. Results from the study demonstrated that metacognitive support in Betty's Brain led to more effective learning of the domain content, as was made evident from the concept map quality measures (Wagster et al., 2007). In analyzing student behaviors and patterns, a classification technique was developed to differentiate high- vs. low-performing students based on the quality of the concept map constructed in relation to the behavior patterns exhibited. It was found that high-performing learners developed a balanced strategy of help-seeking and self-monitoring strategies, while low-performing students would commonly apply a quiz-edit-quiz strategy, where modifications to concept maps were performed only after the system informed the student of an error (Wagster et al., 2007). To further the analysis, it was found that participants in the metacognitively aware Betty's Brain used behaviors indicative of high-performing students when compared against the traditional ITS and Betty's Brain conditions. It was also found that both conditions that involved instructing a teachable agent led to better quality concept maps when compared against those taught by the ITS agent. In terms of transfer, analysis associated with the PFL study showed those individuals in the metacognitive Betty's Brain continued to use effective behavior patterns on the subsequent assignment where scaffolds and feedback functions were removed. This supports that metacognitive and monitoring strategies can transfer across domains and environments, and help students prepare for future learning events (Tan et al., 2007).

For the purpose of this review, it is found that the application of teachable agents in educational technology is a highly effective approach for learning new domains and relationships. In addition, learning by teaching is found to instill effective learning behaviors that aid in producing deep conceptual understanding of a topic through monitoring and regulating procedures. In the studies involving Betty's Brain, the domain is well defined and the learning environment is well controlled. While the task itself is open-ended, the learning environment is designed to support reasoning procedures that equate interactions and solutions with desired outcomes. The caveat with this approach is that a system must provide capabilities for representing knowledge and skill within a format that a computerized agent can make sense of. This limits the types of domains and tasks this approach supports. However, if a domain and learning environment supports teachable agent techniques, this type of instructional design strategy has shown benefit in performance and retention outcomes, as well as in improved metacognitive understanding. A difficult challenge for a wide adoption of this technique is to generalize its application across different types of learning events and interactions. As this learning strategy is tied explicitly to a domain, identifying the elements and variables that translate across problem spaces can provide a conceptual framework for authoring such ITS instances.

Discussion

The studies presented in this review highlight how ITSs can effectively support students in exercising SRL and metacognition during learning. ITSs are unique from other forms of CBLEs in that they have the ability to monitor and assess student performance in real time and provide students with tailored instructional supports that can target the development of SRL behaviors. In this literature review, we identified common strategies delineated from several empirical studies that can translate into high-level instruction-

al strategy instantiations for supporting SRL. Table 1 provides a review of some of the most promising techniques. The first column lists the learning strategy. The second column briefly summarizes the role each strategy plays in the learning process as described above. The third column presents the general effect the strategy was found to have on dependent measures of performance and transfer.

With an established set of strategies found to influence SRL, the next step is to determine if and how each can be implemented within a domain-independent authoring environment such as GIFT. While the strategies listed in Table 1 can be described in generalized terms, the execution of each is dependent to the context of the learning situation and requires tools and methods to link specific interactions to behaviors being monitored by the system. As a result, metacognitively aware ITSs require applying modeled representations of metacognitive behaviors within a domain-specific context. Prompting a learner to perform metacognitive strategies is trivial; the difficulty is determining if the prompt had the intended effect.

In addition, many of the aforementioned strategies are tested in limited cases, giving little evidence of their extensibility across different domains and variations in learning environments. While research has shown the adequacy of these techniques in helping students acquire domain knowledge, there is less evidence to suggest these techniques can truly develop better life-long learners. The shortage of research examining the transfer of metacognitive skills to new topics is concerning. Likewise, there is a shortage of research examining the transfer of metacognitive skills to additional CBLEs. Determining the utility of metacognitive strategy implementations across multiple domains can aid is supporting ITS developers in determining the appropriate interventions to incorporate based on characteristics of the content being taught and the environment being used to do so.

Table 1. Summary of metacognitive strategy types and their effect on learning.

Strategy Type	Its Role in the Learning Process	General Effectiveness
Goal-Setting and Strategic Planning Tools	• Formalizes problem into a hierarchy of goals to plan around, assists in activating prior knowledge, generates initial course of action for achieving defined goals. • Established goals serve as criteria for assessing performance outcomes against. • From the ITS perspective, these metacognitive strategies are generally off-line practices, in that they do not produce data inputs for run-time performance models, making it difficult to assess capability and provide focused instruction on their application.	• Just-in-time training of SRL-based strategies prior to a learning event with hypermedia showed improved performance and mental model shifts when compared against students receiving no upfront training. • Aids used to define goals and sub-goals for a learning space were found to have no effect on an individual's performance when compared to students who received no material highlighting goal structures for the problem space. • TVS was found to improve an individual's ability to solve problems by defining goal targets linked to principal applications of a problem space. The research supports the TVS as an aptitude-treatment interaction based on an individual's potential ability within a domain (Chi & VanLehn, 2010).

Table 1. Summary of metacognitive strategy types and their effect on learning (continued).

Strategy Type	Its Role in the Learning Process	General Effectiveness
Self-Assessment Prompts	• Used to elicit an individual's subjective rating of their ability for a problem space. • The goal of such an intervention is to help students become more aware of their relative strengths and weakness of their knowledge in relation to a learning task.	• Learners were found to improve their ability to identify their strengths while working with the self-assessment tutor developed by Roll et al (2011). • Students transferred the improved self-assessment skills to unsupported problems in the same tutoring environment. • Results are promising and show that relatively simple interventions in an ITS interface can help students become more aware of their strengths with regard to problem solving.
Self-Explanation Prompts	• Prompt students to self-explain or rationalize how a solution was attained. • Typically modeled in two ways: menu-based prompts and free-text windows; each used to elicit verbal response from learner in order to explain how solution was attained. • Forces self-reflection and confirms or denies a learner's understanding for a problem space.	• Prompting students to explain their answers produces robust learning effects, especially if the ITS provides students with feedback on the adequacy of their explanation. • More research is required to better understand the moderating role of feedback across the different self-explanation prompting techniques (i.e., menu-driven vs. free-text).
Cognitive Tools	• Used to assist a learner in solving a problem with intended benefits of reducing cognitive load as tools support processes that are necessary for performance. • Cognitive tools are used to replicate the externalization of knowledge by providing tools and processes that are inherently used by an expert when solving a problem.	• The incorporation of cognitive tools to support monitoring processes and help-seeking behaviors provides learners with the tools and support required to solve a problem based on expert knowledge associated with a domain. • Cognitive tools are found to be used in distinctively different ways across sets of learners and across differences in experience. • Scaffolding is required to assist a novice in appropriately using a cognitive tool as they are designed to support cognitive load associated with performing metacognitive strategies.
Error-Detection and Self-Correction Practices	• Based on an intelligent novice model that allows students to make a few mistakes before intervening with feedback. • When feedback is provided, it steps students through a process of error identification and error correction as opposed to providing immediate corrective feedback.	• Learners in the intelligent novice condition performed better on both immediate and delayed transfer tasks compared to students in an expert model condition that provided immediate feedback following an error. • Learning curve analyses indicate that students who practice error-detection and self-correction strategies learn at a faster rate compared to students in the expert model condition.

Table 1. Summary of metacognitive strategy types and their effect on learning (continued).

Strategy Type	Its Role in the Learning Process	General Effectiveness
Help-Seeking Techniques	• Used to provide students with feedback on their help-seeking skills while interacting with an ITS. • Goal is to have students internalize appropriate help-seeking behaviors that can be generalized from domain to domain.	• Providing students with feedback on their help-seeking skills improved learning. • Improved help seeking skills were also found to transfer to learning new domain-level content. • Roll et al. (2011) found students who had previously received feedback on their help-seeking behavior spent more time focusing on the tutor's hints and asked for fewer hint levels compared students who did not interact with a help-seeking tutor.
Learning by Teaching	• Learners teach artificially intelligent agents. • Self-directed and open-ended in nature, where a learner explores, integrates, and structures knowledge before they can assist another in doing the same. • Requires active self-direction, and encompasses three aspects of learning interaction: structuring, taking responsibility, and reflecting (Biswas et al., 2005). • Teachable agent can provide metacognitive support by externalizing strategies that are used to learn material.	• Learning by teaching is found to instill effective learning behaviors that aid in producing deep conceptual understanding of a topic through monitoring and regulating procedures. • There is empirical support that metacognitive and monitoring strategies can transfer across domains and environments and help students prepare for future learning events (Tan et al., 2007). • Receiving metacognitive support during learning by teaching event promotes the transfer of effective behaviors when support is removed in subsequent interactions. • System must provide capabilities for representing knowledge and skill within a format that a computerized agent can make sense of.

Applying Metacognitive Support in GIFT

For GIFT to operate outside a laboratory setting as a domain-independent authoring environment, supports must be included for helping system developers establish these metacognitive tutoring practices. Currently, GIFT is designed to support adaptive authoring by recommending pedagogical strategies to an instructional designer, who is then responsible for translating that strategy into a tactic for execution in the learning environment. These strategies are currently maintained for domain-level learning and vary the type of guidance based on generalized timing and specificity dimensions. The caveat is that, because GIFT is domain agnostic, all of these processes need to be defined in a generalized fashion so that they extend across domain implementations. GIFT monitors performance through an ontological representation of a domain by expressing objectives and concepts in a relational hierarchy. For each concept identified in the hierarchy, an assessment is authored that designates metrics linked to competency. These metrics are used to produce a learner state for each defined concept, which is used by the pedagogical model to inform guidance functions. From there, GIFT makes informed pedagogical recommendations on a domain-independent level (e.g., provide hint, provide prompt), leaving it to the instructor to author that strategy as an actionable tactic within the training environment (Goldberg et al., 2012). An example would be GIFT requesting a hint for concept 2.1.2, with the tactic linked to that hint being the specific prompt to display.

In this instance, a system developer is required to author multiple levels of tactics for each concept being tracked within the domain model. Having multiple instructional tactics linked to a specific concept enables the system to vary the level of detail provided in feedback messages based on individual differences associated with a learner. In the event that a system requires updates to task procedures and assessments, tactic definitions for each affected concept will need to be updated as well. This can be a taxing process on the course administrator if the task being trained modifies on a regular basis, which is frequent within industry- and military-based settings. When it comes to feedback linked to SRL and metacognition, what are the implications with respect to the current approach?

For GIFT to support SRL-based tutoring practices, there needs to be a set of defined persistent metacognitive strategies that can be applied across domain applications, such as those identified throughout this review. Furthermore, to support the already existing schema GIFT operates around, a set of desired SRL- and metacognitive-based behaviors must be established in the domain model (i.e., help-seeking skill, self-explanation skill). This explicitly defines the behaviors a system is going to monitor, assuming there are modeling approaches to build assessment rules around. These established behaviors and assessments in the domain model can be used to inform pedagogical interventions intended to influence a learner's interaction within the environment. Just as described above, specific tactics must be authored for each SRL-based concept tracked when a feedback request is received from the pedagogical model. However, the use of metacognitive feedback prompts can be expressed in a general enough fashion, so as to extend beyond being a strategy recommendation alone through the use of standard reflective and self-assessment prompts. When GIFT requests a self-reflection prompt, a generic tactic can be delivered that states "can you please explain how you reached your solution." This type of prompt is void of domain-relevant information and can be extended to any type of ITS supporting self-reflection. The important aspect of this type of intervention is to assess a learner's response to the prompt, either through free-response or based on a selection from a bank of choices. Regardless of the approach, determining the effect a tactic has on metacognitive behavior is dependent on the context of the learning event. To further explain this concept, the schematic view of the help-seeking model presented in Koedinger et al. (2009) was adapted based on current functions and message structures in GIFT (see Figure 2). The adapted model is potentially useful to GIFT development as it outlines the sequence of behavioral and mental processes a learner engages in when solving a set of problems.
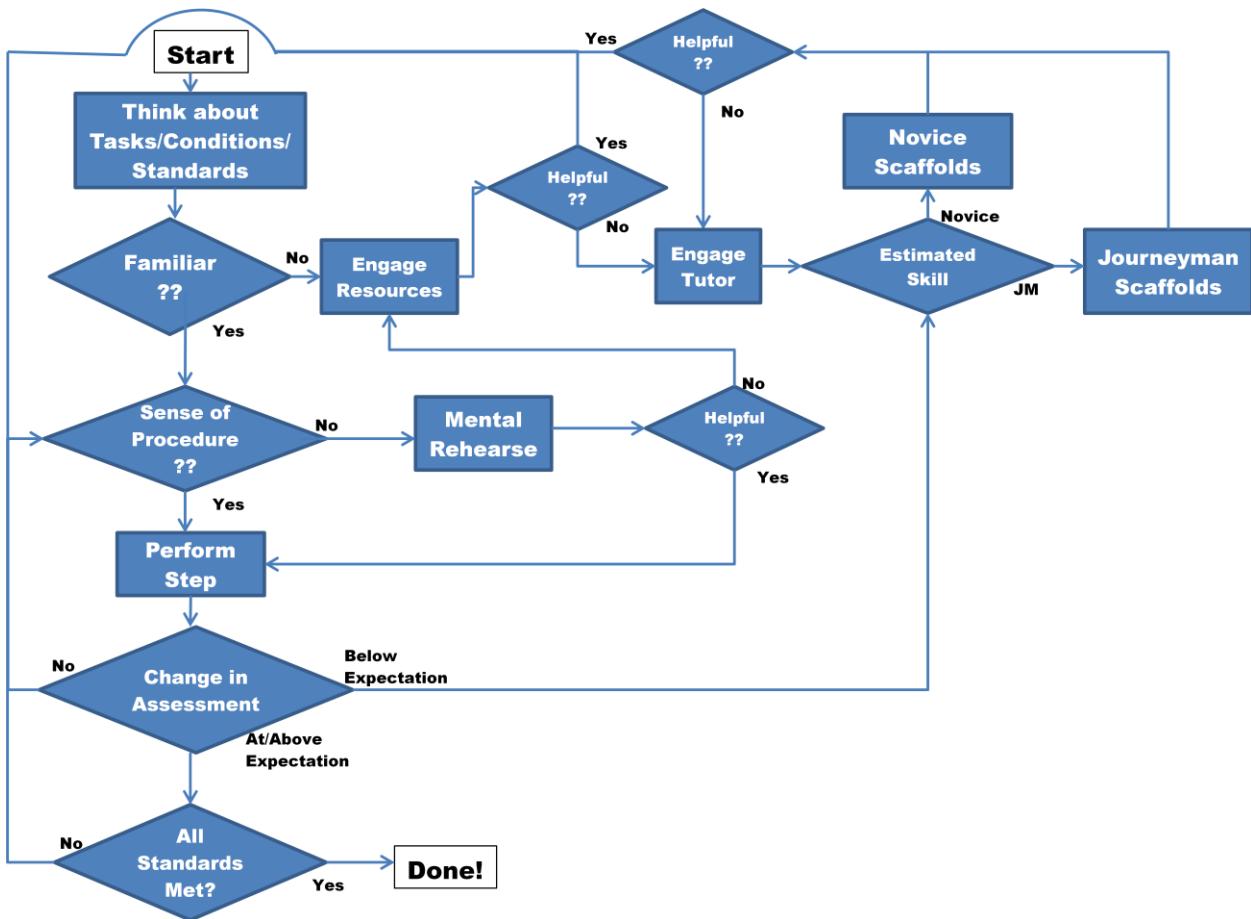
Figure 2. Adapted help-seeking model based on GIFT interactions.

The model shows aspects of tutor engagement based on student-initiated and system-initiated activators. In addition, this model highlights a path of preferred interaction in terms of sequence, where a learner is expected to engage in provided help resources before they rely on the tutor for assistance. When solving a set of problems, student-initiated tutor engagement is triggered by a student, with the assumption that a learner is at an impasse with no notion for how to proceed. Reactive engagement is based on performance outcomes when a set of actions results in an explicit performance state change reported by the learner model. This state change designation corresponds with evidence in the environment that the learner is below expectation on a concept, where a pedagogical intervention is delivered to assist in correcting an error or misconception. If a system is able to monitor and track interactions and patterns of interactions, a model can be developed to gauge how a learner is using the resources the ITS provides, and if certain metacognitive behaviors are being ignored, such as seeking help to identify the next step to apply in solving a problem. While this model is adapted on the conceptual level, we believe a generalized approach can be implemented that can track a user's interaction and make inference on help-seeking practices and possible gaming behaviors (Baker et al., 2013).

An important component of this adapted model is determining the functions made available to a learner when they initiate engagement with the tutor themselves. What is the role of the tutor in this capacity? What information is required to select metacognitive tutoring practices over domain-relevant guidance? If the system can determine whether the learner engaged the tutor prior to executing any help-seeking behaviors, does the system prompt the learner to use the available resources before asking for help? These

types of questions are of interest to the GIFT developers, as heuristics will need to be identified that guide these types of pedagogical decisions. In the context of system-initiated tutor engagement, the adapted model incorporates scaffolds based largely on an individual's estimated skill level, as deemed by assessed prior knowledge and metrics associated with real-time performance. These scaffolds highlight the requirement of sub-models within this schematic layout, where other variables linked to individual differences may dictate the timing and specificity of guidance. While this model fleshes out the behaviors and actions of an ideal student, how to proceed pedagogically when tutor engagement is enacted is an open research question.

With that said, metacognitive tutoring practices are inherently linked to the domain they are applied within. The same can be said for standard pedagogical strategies on the domain level. While you can define fixed recommendations, the information used to trigger those strategies is tied directly to domain-level representations of actions and state assessments. Based on these assumptions, identifying and authoring metacognitive strategies to implement is not complex; however, triggering a strategy and assessing the impact an executed tactic has on subsequent performance is dependent on the system's ability to link user interactions to defined SRL and metacognitive behaviors present in the domain ontology representation. This brings to light the significant challenge associated with this form of peda-gogical intervention, which is establishing formalized modeling methods that can accurately gauge an individual's ability at applying and regulating efficient learning behaviors across multiple domains and learning environments. Before modeling techniques can be researched, a set of metacognitive behaviors must be established that warrant real-time monitoring for informing intervention. Using the methods reviewed above is a good starting place for determining what information is required to assess those behaviors and skills, and if they can be represented in a format that is compatible with GIFT functions.

Conclusion

In this chapter, we reviewed the theoretical construct of SRL and the role metacognitive processes play when an individual is learning and solving problems on their own. While ITSs apply modeling techniques to assist an individual in regulating interaction, there is a push by the learning sciences community to use these technologies to help develop valuable skills that increase an individual's problem-solving ability. As the SRL model identifies an array of strategies that an ideal student employs, ITS developers are exploring innovative ways to direct and influence a learner in enacting these behaviors, with the overall objective of instilling these skills for future use in similar settings. Significant progress has been made over the past decade in terms of ITS functions and processes supporting metacognitive tutoring. Yet, there is still much research left to optimize the use of educational technology in developing effective life-long learners. Using GIFT as an experimental testbed to investigate the utility of these strategies across multiple use cases can help accelerate the progression of metacognitive tutoring methodologies.

References

Aleven, V. & Koedinger, K. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*(2), 147-179.

Aleven, V. & Koedinger, K. R. (2000). *The need for tutorial dialog to support self-explanation.* Paper presented at the Building dialogue systems for tutorial applications, papers of the 2000 AAAI Fall Symposium.

Aleven, V., Mclaren, B., Roll, I. & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education, 16*(2), 101-128.

Aleven, V., Ogan, A., Popescu, O., Torrey, C. & Koedinger, K. (2004). *Evaluating the effectiveness of a tutorial dialogue system for self-explanation.* Paper presented at the Intelligent Tutoring Systems.

Anderman, E. & Corno, L. (2013). *Handbook of educational psychology*: Routledge.

Atkinson, R. K., Renkl, A. & Merrill, M. M. (2003). Transitioning From Studying Examples to Solving Problems: Effects of Self-Explanation Prompts and Fading Worked-Out Steps. *Journal of Educational Psychology, 95*(4), 774.

Azevedo, R. (2009). Theoretical, methodological, and analytical challenges in the research on metacognition and self-regulation: A commentary. *Metacognition & Learning, 4*(1), 87-95.

Azevedo, R. & Aleven, V. A. (2013). *International handbook of metacognition and learning technologies* (Vol. 26): Springer.

Azevedo, R. & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*(3), 523.

Azevedo, R., Cromley, J. G. & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology, 29*(3), 344-370.

Bandura, A. (1991). *Self-regulation of motivation through anticipatory and self-reactive mechanisms.* Paper presented at the Perspectives on motivation: Nebraska symposium on motivation.

Bargh, J. A. & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72*(5), 593.

Biswas, G., Leelawong, K., Belynne, K., Viswanath, K., Vye, N., Schwartz, D. & Davis, J. (2004). *Incorporating self regulated learning techniques into learning by teaching environments.* Paper presented at the The Twenty Sixth Annual Meeting of the Cognitive Science Society.

Biswas, G., Leelawong, K., Schwartz, D. & Vye, N. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence, 19*(3-4), 363-392.

Biswas, G., Schwartz, D. & Bransford, J. (2001). *Technology support for complex problem solving: From SAD environments to AI.* Paper presented at the Smart machines in education.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A., Dunlosky, J. & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417-444.

Bransford, J., Barron, B., Pea, R. D., Meltzoff, A., Kuhl, P., Bell, P., . . . Reeves, B. (2005). Foundations and opportunities for an interdisciplinary science of learning. *The Cambridge handbook of the learning sciences*, 39-77.

Bransford, J. D., Brown, A. L. & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington, D.C.: National Academy Press.

Brown, J. S., Collins, A. & Newman, S. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. *Knowing, learning, and instruction: Essays in honor of Robert Glaser, 487*.

Butler, D. L., Cartier, S. C., Schnellert, L., Gagnon, F. & Giammarino, M. (2011). Secondary students' self-regulated engagement in reading: Researching self-regulation as situated in context. *Psychological Test and Assessment Modeling, 53*(1), 73-105.

Chi, M. & VanLehn, K. (2010). Meta-Cognitive Strategy Instruction in Intelligent Tutoring Systems: How, When, and Why. *Educational Technology and Society, 13*(1), 25-39.

Conati, C. (2009). *Intelligent Tutoring Systems: New Challenges and Directions.* Paper presented at the IJCAI.

Conati, C. & Vanlehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education (IJAIED), 11*, 389-415.

Corbett, A. T. & Anderson, J. R. (2001). *Locus of Feedback Control in Computer-Based Tutoring: Impact on Learning Rate, Achievement and Attitudes*. Paper presented at the ACM CHI'2001 Conference on Human Factors in Computing Systems, New York.

d Baker, R. S., Corbett, A., Roll, I., Koedinger, K., Aleven, V., Cocea, M., . . . Mathews, M. (2013). Modeling and studying gaming the system with educational data mining *International Handbook of Metacognition and Learning Technologies* (pp. 97-115): Springer.

Flavell, J. H., Miller, P. H. & Miller, S. A. (1985). *Cognitive development*: Prentice-Hall Englewood Cliffs, NJ.

Glaser, R., Chi, M. T. & Farr, M. J. (1988). *The nature of expertise*: Lawrence Erlbaum Associates.

Goldberg, B., Brawner, K. W., Sottilare, R., Tarr, R., Billings, D. R. & Malone, N. (2012). *Use of Evidence-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) 2012, Orlando, FL.

Hadwin, A. F. & Winne, P. H. (2001). CoNoteS2: A software tool for promoting self-regulation. *Educational Research and Evaluation, 7*(2-3), 313-334.

Johnson, C. I. & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment *Computers in Human Behavior, 26*(6), 1246-1252.

Karabenick, S. A. & Newman, R. S. (2006). *Help seeking in academic settings: Goals, groups, and contexts*: Psychology Press.

Koedinger, K., Aleven, V., Roll, I. & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. *Handbook of metacognition in education*, 897-964.

Lajoie, S. P. (2008). Metacognition, self regulation, and self-regulated learning: a rose by any other name? *Educational Psychology Review, 20*(4), 469-475.

Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from avionics and medicine. *Development of professional expertise: Toward measurement of expert performance and design of optimal learning environments*, 61-83.

Leelawong, K. & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education, 18*(3), 181-208.

Mathan, S. A. & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist, 40*(4), 257-265.

McCurdy, N., Naismith, L. & Lajoie, S. (2010). *Using Metacognitive Tools to Scaffold Medical Students Developing Clinical Reasoning Skills.* Paper presented at the Cognitive and Metacognitive Educational Systems: Papers from the AAAI Fall Symposium.

Nathan, M. J. (1998). Knowledge and situational feedback in a learning environment for algebra story problem solving. *Interactive Learning Environments, 5*(1), 135-159.

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into practice, 41*(4), 219-225.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-407.

Pintrich, P. R. & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33.

Poitras, E., Lajoie, S. & Hong, Y.-J. (2012). The design of technology-rich learning environments as metacognitive tools in history education. *Instructional science, 40*(6), 1033-1061.

Poitras, E. G. & Lajoie, S. P. (2013). A domain-specific account of self-regulated learning: the cognitive and metacognitive activities involved in learning through historical inquiry. *Metacognition and Learning, 8*(3), 213-234.

Roll, I., Aleven, V., McLaren, B. M. & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*(2), 267-280.

Roll, I., Holmes, N. G., Day, J. & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. *Instructional science, 40*(4), 691-710.

Schraw, G., Crippen, K. J. & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*(1-2), 111-139.

Schunk, D. H. & Schunk, D. (2005). Self-Regulated Learning: The Educational Legacy of Paul R. Pintrich. *Psychologist, 40*, 85-94.

Tan, J., Biswas, G. & Schwartz, D. (2006). *Feedback for metacognitive support in learning by teaching environments.* Paper presented at the Proceedings of the 28th Annual Meeting of the Cognitive Science Society, Vancouver, Canada.

Tan, J., Wagster, J., Wu, Y. & Biswas, G. (2007). Effect of metacognitive support on student behaviors in learning by teaching environments. *Frontiers In Artificial Intelligence and Applications, 158*, 650.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist, 46*(4), 197-221.

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., . . . Wintersgill, M. (2005). The Andes Physics Tutoring System: Lessons Learned. *International Journal of Artificial Intelligence in Education, 15*(3), 1-47.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T. & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction, 21*(3), 209-249.

Wagster, J., Tan, J., Wu, Y., Biswas, G. & Schwartz, D. (2007). *Do learning by teaching environments with metacognitive support help students develop better learning behaviors.* Paper presented at the The twenty-ninth Annual Meeting of the Cognitive Science Society, Nashville, Tennessee.

Weerasinghe, A., Mitrovic, A. & Martin, B. (2007). *Towards a General Model for Supporting Explanations to Enhance Learning.* Paper presented at the Proceedings of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work.

Weerasinghe, A., Mitrovic, A. & Martin, B. (2009). Towards individualized dialogue support for ill-defined domains. *International Journal of Artificial Intelligence in Education, 19*(4), 357-379.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. *Self-regulated learning and academic achievement: Theoretical perspectives, 2*, 153-189.

Winne, P. H. & Hadwin, A. F. (1998). Studying as self-regulated learning.

Zimmerman, B. J. (2002). Achieving academic excellence: A self-regulatory perspective. *The pursuit of excellence through education*, 85-110.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Zimmerman, B. J. & Campillo, M. (2003). Motivating self-regulated problem solvers. *The psychology of problem solving*, 233-262.

Zimmerman, B. J. & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. *Handbook of metacognition in education*, 299-315.

Zimmerman, B. J. & Schunk, D. H. (2011). *Handbook of self-regulation of learning and performance*: Taylor & Francis.

CHAPTER 11 – A Combined Theory- and Data-Driven Approach for Interpreting Learners' Metacognitive Behaviors
in Open-Ended Tutoring Environments

Gautam Biswas, James R. Segedy, and John S. Kinnebrew
Vanderbilt University

Introduction

Adapting to learners' needs and providing useful individualized feedback to help them succeed has been a hallmark of most ITSs (e.g., Park & Lee, 2004). These systems take explicit actions (Puntambekar & Hubscher, 2005), such as reminding learners of relevant information or modifying the learning activity to support learning processes (Lajoie & Azevedo, 2006; Segedy, Loretz & Biswas, 2013). To promote deep learning, critical thinking, and problem-solving skills in science, technology, engineering, and mathematics (STEM) disciplines, researchers have begun developing open-ended learning environments (OELEs) that provide a learning context and a set of tools for learning and solving complex problems (Land, 2000). To be successful in these environments, learners have to employ metacognitive processes to manage, coordinate, and reflect on relevant cognitive processes as they search for and interpret information, and apply it to construct and test potential problem solutions. This can present significant challenges to novice learners. They may lack the proficiency to use the system's tools, and the experience and understanding necessary to explicitly regulate their learning and problem solving. Traditionally, learning behavior in intelligent tutors and OELEs are assessed with theory-driven metrics and context-driven hypotheses about the students' learning tasks. In recent years, data-mining techniques that analyze students' logged activity data have been used to discover important aspects of how students learn (Romero & Ventura, 2009).

This chapter discusses a framework for analyzing learning activity data in OELEs that combines top-down metrics and bottom-up pattern discovery. This integrated framework can be employed to build detailed models of students' learning behaviors and strategies, and subsequently to identify opportunities for providing adaptive scaffolds to students as they use the system. For top-down, theory-driven analysis of learning behaviors, our framework focuses on 1) the learner's acquisition and application of knowledge in the OELE and 2) the impact of these activities on the learning task (e.g., whether an action results in progress toward task completion (Segedy, Biswas & Sulcer, 2014)). For bottom-up, data-driven discovery of learning behaviors, our framework employs data-mining techniques for identifying frequent and important action patterns from logs of student activity. Our approach enhances the analysis and assessment of student learning behavior by combining the complementary top-down and bottom-up techniques. This allows us to identify specific learning behaviors for a group of students, behavior differences between groups that are relevant to understanding their approach to learning in the environment, and the connections between specific patterns of activity and the relevant skills or strategies for learning and problem solving.

The top-down, theory-driven metrics are used for evaluating and differentiating instances of patterns discovered in the data to better understand whether or not the discovered patterns were used as part of coherent strategies and, if so, which ones. The theoretical measures also provide valuable information about individual differences among students that may employ the same pattern of actions but in different manners or for different purposes. We instantiate this task-driven analytic framework in the context of Betty's Brain (Leelawong & Biswas, 2008), an OELE where students learn science by constructing causal

models. A case study illustrates the benefits of incorporating top-down and bottom-up techniques in concert to characterize the learning behavior of students in an OELE.

Background: Metacognition

Flavell (1979) defined metacognition as "*thinking about one's own thinking*." From an information-processing perspective, Winne (1996) described cognition as dealing with knowledge of objects and operations on objects (the object level), while characterizing metacognition as the corresponding meta level that contains information about when to use particular cognitive processes and how to combine them to accomplish larger tasks. Metacognitive monitoring brings the two levels together, as it describes the process of observing and evaluating one's own execution of cognitive processes in order to exercise control for improving cognition. When applied to learning situations, metacognition encompasses (Hennessey, 1999; Martinez, 2006):

- The knowledge and control learners exhibit over their thinking and learning activities;

- Awareness of one's own thinking and conceptions;

- Active monitoring of one's cognitive processes;

- An attempt to control and regulate one's cognitive processes to support learning; and

- The application of heuristics or strategies for developing one's own approach to solving problems.

In general, control or regulation of cognition (Brown et al., 1983), and application of strategies to regulate one's learning are fundamental components of metacognition. Winne and Hadwin (1998; 2008) have proposed a model of self-regulated learning called COPES. Learning according to this model occurs in four weakly sequenced and recursive stages: 1) *task definition*, where the students develop their own understanding of the learning task, 2) *goal setting and planning*, which follow the task definition phase and represent the students' approach to working on the learning task, 3) *enactment of tactics*, which represents that phase where the students' carry out their plans for learning, and 4) *adaptations to meta-cognition*, which are linked to both in-the-moment adjustments of one's tactics and post-hoc evaluation of one's approach based on successes and failures achieved during enactment. Like COPES, we adopt a *task modeling approach* to interpret students' learning behaviors in the Betty's Brain environment. Patterns derived from students' activity sequences can be interpreted as cognitive and metacognitive processes associated with the learning tasks. This approach emerges from the link between cognitive skill proficiency and metacognitive planning (Land, 2000; Veenman, 2012). Metacognitive knowledge by itself may not be sufficient to achieve success, especially when learners lack the cognitive skills and background knowledge necessary for understanding and organizing the problem under study (Bransford, Brown & Cocking, 2000). We take this into account by incorporating and linking both cognition and metacognition in the task model employed during the analysis of students' learning behaviors.

Related Research

Several OELEs have been designed to provide adaptive scaffolds. For example, Ecolab (Luckin & Hammerton, 2002) intervenes whenever the student specifies an incorrect relationship (e.g., caterpillars eat thistles). It notifies students that the relationship is incorrect and provides corrective hints. Should students continue to struggle, the system will tell students exactly how to complete the task (e.g., you

need to model the relationship "*caterpillars eat grass*"). Learners using Ecolab are free to choose the order in which they perform their learning activities, and the system uses information about the number of student errors to select activities that are within the student's ZPD (Luckin & du Boulay, 1999). If students choose a learning activity that the system has deemed too easy or too difficult, the system prompts them to reconsider their choice. In Crystal Island (Spires, Rowe, Mott & Lester, 2011), learners take on the role of a microbiologist to find the identity and source of an infectious disease plaguing the research station. As learners explore the island and complete tasks, the system keeps track of the number of laboratory experiments that learners have conducted, and after every five experiments, it intervenes and requires students to correctly answer questions about microbiology. The agent also tracks information that learners encounter while conversing with computer-controlled characters, and it quizzes students on that information later.

These two analysis techniques focus on either 1) the correctness of student actions (as in Ecolab) or 2) counts of student actions (as in Crystal Island). However, these approaches do not characterize how students *coordinate their use of system tools* to complete their learning tasks. Our approach combines a cognitive and metacognitive model of the learning task with theory-driven measures to analyze students' activities and use of tools that are directed to acquire information and subsequently apply it to complete their learning and problem-solving tasks. However, an analysis based on these theory-driven measures and patterns of actions derived from the task model may not be sufficient to cover the wide variety of different behaviors and strategies students employ during learning and problem solving. Therefore, our analysis framework includes data-driven sequence mining techniques to identify the patterns of activity that students actually employ in the learning environment.

Researchers have applied sequence mining techniques to educational data to better understand learning behaviors. Perera et al. (2009) provide mirroring and feedback tools to support effective teamwork among students collaborating on software development using an open-source professional development environment called Trac[1]. Their analysis used sequence mining to derive students' learning behaviors, and Perera showed that mirroring and feedback helped all groups improve their work by emulating the behaviors of the strong groups. In previous work, we have compared sequential patterns derived from student activity sequences to identify ones that differ in use between two or more groups of students (Kinnebrew, Loretz & Biswas, 2013) and over time (Kinnebrew, Mack & Biswas, 2013). Nesbit et al. (2007) use sequential pattern mining to find the longest common subsequences across a set of action files from the *gStudy* learning environment to study how students self-regulate as they learn. Other researchers have also employed sequential pattern mining to generate student models for customizing learning to individual students (Amershi & Conati, 2009; Tang & McCalla, 2002).

To identify general behavior patterns that are common across students, sequence mining techniques generally employ a high-level description of student actions (e.g., the student read a page for a long period of time). However, a single behavior pattern (defined by a short sequence of these high-level action descriptions) could be interpreted in multiple ways, depending on the context and relationship of an action (e.g., exactly what information could be gained from the page read) with other prior or subsequent actions (e.g., the student next edited the causal map, adding information acquired from reading the page). In this chapter, we explicitly map the high-level patterns of student actions back into the context of students' complete sequences of activities, employing the additional, specific details about each action and surrounding actions to calculate theory-driven measures that differentiate behaviors that would otherwise be represented by the same high-level pattern of actions. This contextualization and application of theory-driven measures allows us to link specific instances of an activity pattern to the relevant skills or strategies in the cognitive/metacognitive task model. Further, it allows us to identify additional strate-gies (both effective and ineffective) that are employed by students in order to extend the cogni-

---

[1] http://trac.edgewall.org/

tive/metacognitive task model for a more complete description of the domain and more comprehensive learner models.

Betty's Brain

The Betty's Brain learning environment (Leelawong & Biswas, 2008) presents students with the task of teaching a virtual agent, Betty. Students teach Betty a science topic by constructing a visual causal map that represents the relevant science phenomena as a set of entities connected by directed links that represent causal relations. Once taught, Betty can use the map to answer causal questions and explain those answers. The goal for students using Betty's Brain is to teach Betty a causal map that matches a hidden, expert model of the domain. The students' learning and teaching tasks are organized around three activities: 1) reading hypertext resources, 2) building the map, and 3) assessing the correctness of the map. The hypertext resources describe the science topic under study (e.g., climate change) by breaking it down into a set of sub-topics. Each sub-topic describes a system or a process (e.g., the greenhouse effect) in terms of entities (e.g., absorbed heat energy) and causal relations among those entities (absorbed heat energy increases the average global temperature). As students read, they need to identify causal relations and then explicitly teach those relations to Betty by adding them correctly to the current causal map. Figure 1 illustrates the Betty's Brain system interface.
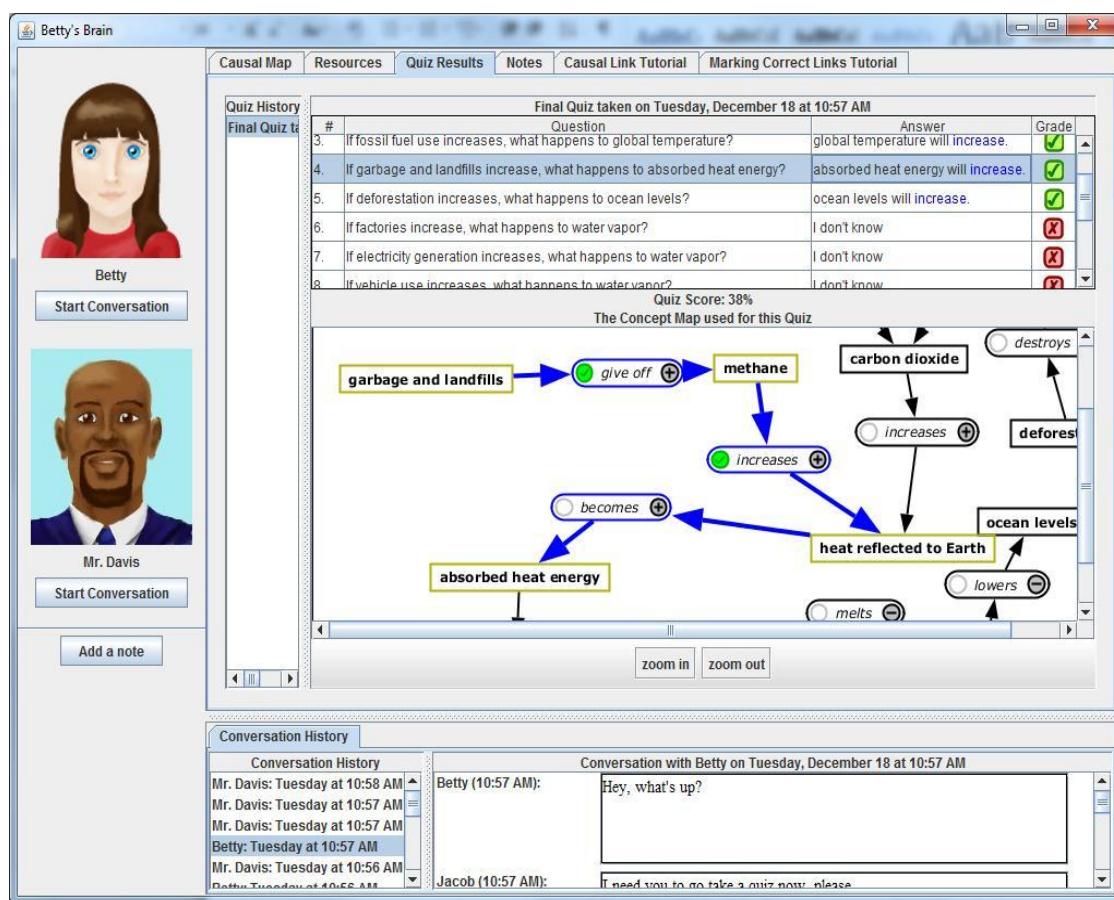
Figure 1. Betty's Brain system showing the quiz interface.

Learners can assess the quality of their current map in two ways. First, they can ask Betty to answer a cause-and-effect question using a template. After Betty answers the question, learners can ask Mr. Davis, another pedagogical agent that serves as a mentor, to evaluate her answer. If the portion of the map that Betty uses to answer the question matches the expert model, then Betty's answer is correct. Learners can also have Betty take a quiz on one or all of the sub-topics in the resources. Quiz questions are selected dynamically by comparing Betty's current causal map to the expert map. Since the quiz is designed to reflect the current state of the student's map, a set of questions is chosen (in proportion to the completeness of the map) for which Betty will generate *correct* answers. The rest of the quiz questions produce either *incorrect* or *incomplete* answers. These answers can be used to infer which causal links are correct and which causal links may need to be revised or removed from the map. Should learners be unsure of how to proceed in their learning task, they can ask Mr. Davis for help via a menu-based conversation that allows the user to choose from a set of pre-specified options. Mr. Davis responds by asking learners about what they are trying to do and responds with suggestions appropriate to the user's indicated goals.

Framework Integrating Theory- and Data-Driven Analysis

Our framework for analyzing OELE learning activity data integrates top-down information acquisition/application measures and bottom-up sequential pattern discovery as illustrated in Figure 2. The analysis involves sequential pattern mining to identify common action patterns, mapping identified patterns back into the complete activity sequences to analyze the context and specific details of the actions corresponding to the pattern with theory-driven measures, and linking the analyzed behaviors back to skills and strategies in the cognitive/metacognitive task model. In the following sections, we describe the specific measures and data-mining techniques employed in this framework, and we instantiate it with the cognitive/metacognitive task model for analysis of Betty's Brain data.
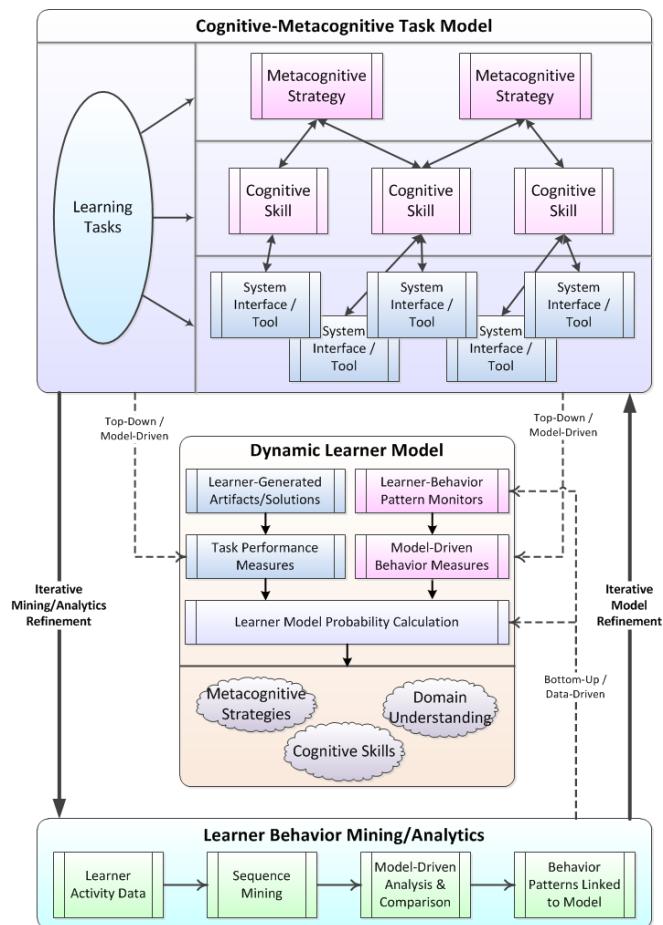
Figure 2. Integrated analytics framework.

Theory-Driven, Top-Down Analysis

The theory-driven portion of our integrated framework, illustrated in Figure 3, incorporates a cognitive and metacognitive model linked to the tasks that students are expected to complete as they progress through an open-ended learning task. In order to analyze Betty's Brain data, we have developed a task model that represents student activities as a set of cognitive and metacognitive activities related to 1) *knowledge construction*, which consists of both *information seeking & acquisition* and *solution construction*; and 2) *solution evaluation* (Kinnebrew, Segedy & Biswas 2014; Segedy, Biswas & Sulcer 2014).[2] The directed links in the model represent dependency relations. The model indicates that each of these high-level characterizations involves a set of metacognitive tasks, and each specific task can be accomplished by applying any of a number of metacognitive strategies. Information-seeking tasks depend on one's ability to read, understand, interpret, and translate information from the resources. Solution construction tasks depend on one's ability to apply information gained during information seeking and solution evaluation to constructing and refining the causal map. Finally, solution evaluation tasks depend on the learner's ability to interpret the results of solution assessments (question evaluations and quizzes) as actionable information that can be used to refine the solution in progress.

---

[2] In this chapter, we do not discuss the goal setting and planning or help-seeking behaviors
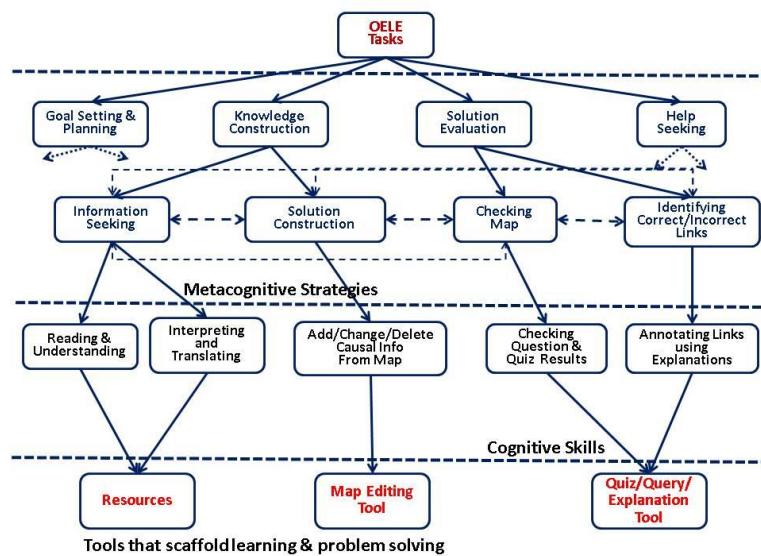
Figure 3. Cognitive/metacognitive task model for Betty's Brain.

The structure of the cognitive and metacognitive task model provides two key pieces of theory-driven information that can be used to judge the quality of student behaviors in Betty's Brain. First, the dependency relations between metacognitive and cognitive tasks indicate that one must use information about the student's cognitive ability levels when analyzing students' behaviors in an OELE. Students who lack the required cognitive abilities are not likely to succeed in applying metacognitive strategies when working on the higher-level task. Second, the dependency of solution construction on information seeking and solution evaluation tasks indicates that students must coordinate their use of system tools in an appropriate manner, to filter information and apply what they learn to the construction of a correct solution. Such coordination requires metacognitive regulation as students decide how to apply the information they have learned. Thus, analyzing students' learning behaviors must also assess students' metacognitive regulation through their ability to logically coordinate their use of multiple tools within the system.

To assess students' cognitive ability levels, our approach judges each action students take on the system in terms of its *effectiveness*. Actions in an OELE are effective if they move the learner closer to their task goal, and students with higher proportions of effective actions are considered to have a higher mastery of the cognitive processes listed in the model. In this chapter, we focus on *solution construction effectiveness*. Solution construction actions are effective when they improve the overall quality of the solution in progress.

To assess one aspect of student metacognitive regulation, our approach evaluates student behaviors using a measure of coherence called *action support*. Support for a particular student action represents the extent to which it is informed by information that could be acquired through previous actions. For example, information-seeking actions (e.g., reading about a causal relationship) can provide support for future solution construction actions (e.g., adding the corresponding causal link to the map). Students with higher proportions of supported actions are considered to have a higher mastery of strategies for coordinating their use of tools within the environment.

Data-Driven, Bottom-Up Analysis

To identify student behaviors in the learning environment, our framework applies a sequential pattern mining algorithm to logged records of student actions. To effectively perform sequential data mining on learning interaction traces, raw logs must first be transformed into an appropriate sequence of actions. In this step, researcher-identified categories of actions, corresponding to the relevant system tools and interfaces in the cognitive/metacognitive task model, define the set of actions that may appear in the activity sequences. This filters out irrelevant information (e.g., cursor position) and combines qualitatively similar actions (e.g., performing the same action through different interface features). The resulting patterns are then input to a sequential pattern mining algorithm. In the analysis presented, we employ an algorithm (from Pex-SPAM (Ho, Lukov & Chawla, 2005)) to identify patterns that meet a given sequence mining support threshold, i.e., the identified patterns occur in at least a given percentage of the sequences. To identify patterns that are common to the majority of the students, we apply a sequence mining support threshold of 50% on the sequential pattern mining algorithm.

Integrating Theory-Driven Measures with Data-Driven Analysis

Common behavior patterns identified by the sequence mining algorithm have to be interpreted and analyzed by researchers to identify a relevant subset of important patterns that provide a basis for generating actionable insights (e.g., how to scaffold user interactions with the learning environment to encourage specific, productive behaviors). Our framework maps the patterns back into student sequences to identify the individual occurrences of each pattern and then analyzes these instances of the patterns in context to more effectively interpret and differentiate different behaviors that result in the same action pattern. For example, the sequential pattern mining algorithm might identify the pattern "A brief read followed by adding a correct causal link." This pattern leaves out some detailed information, such as the specific page read and particular causal link added to the map. Therefore, this pattern of brief reading and adding a causal link may happen many different times, even for a single student, but involve different pages and causal links. Further, such information about each specific instance of the pattern is necessary for determining whether the page read discusses the concepts and their relationship that were represented in the causal link added. To differentiate these distinct instances of each pattern, we employ the information acquisition and application measures along with a measure of *pattern coherence*, which describes whether or not actions in a specific instance of a pattern are such that 1) an earlier action provides action support for a later action or 2) two actions in the pattern are supported by a common previous action (which may have occurred before the pattern instance).

By taking into account the action support and effectiveness of the discovered frequent pattern instances, our framework can distinguish between effective and ineffective behaviors that are defined by the same action pattern. The support and effectiveness measures apply to individual actions, and may be used to refine the definition of canonical actions by applying thresholds to the action support and effectiveness values. For example, this may result in further classifying a read statement as an *ineffective-read* versus an *effective-read*. Whereas this information may be very useful in contextualizing the meaning and use of derived patterns that contain these actions, they may also have the effect of reducing the frequency of the observed pattern. For example, the qualification of actions by their action support and effectiveness measures may reduce the occurrence of patterns that contain these actions to below 50%, making those patterns ineligible for further analysis. To overcome this problem, our integrated framework incorporates these measures for further interpretation only after discovering common patterns using the sequence mining approach.

OELE Study and Results

Our analysis is based on data collected from a recent middle school classroom study with Betty's Brain. The study tested the effectiveness of two support modules designed to scaffold students' understanding of cognitive skills and metacognitive strategies important for success in building the correct causal map. The *Knowledge Construction* (KC) support module scaffolded students' understanding of how to construct knowledge by identifying causal relations in the resources, and the *Solution Evaluation* (SE) support module scaffolded students' understanding of how to monitor Betty's progress using the quiz results to identify correct and incorrect causal links on Betty's map. Participants were divided into four treatment groups. The *Knowledge Construction group* (KC-G) used a version of Betty's Brain that included the KC support module and a causal link tutorial that they could access at any time during learning. The *Solution Evaluation group* (SE-G) used a version of Betty's Brain that included the SE support Module and a marking links correct tutorial that they could access at any time during learning. In addition to the KC and SE groups, the experiment included a *Control group* (Con-G) and a *Full Support group* (Full-G). The control group used a version of Betty's Brain that included neither the tutorials nor the support modules, and the full support group used a version of Betty's Brain that included both tutorials and support modules.

Students used the Betty's Brain system to learn about climate change. The expert map includes 22 concepts and 25 links representing the *greenhouse effect*, *human activities* linked to the greenhouse effect, and potential impact of the greenhouse *effect on climate*. The hypermedia resources on climate change contain 31 hypertext pages with a Flesch-Kincaid reading grade level of 8.4. Learning was assessed using a pre-post test design. Each written test was made up of five questions that asked students to consider a given scenario (e.g., a significant increase in the use of road vehicles) and explain its causal impact on climate change. The maximum combined score for the five questions was 16.

The experimental analysis reported in this paper used data from 20 KC-G students, 17 SE-G students, 15 Con-G students, and 16 Full-G students. The study was conducted for 9 school days, with students participating for a 60-minute class period each day. The first four class periods included a pre-test and training with Betty's Brain and causal modeling. Students then spent four class periods (days 5-8) working with their respective versions of the Betty's Brain system with minimal intervention by the teachers and the researchers. On the ninth day, students completed the post-test that was identical to the pre-test.

Log Analysis

To extract the activity sequences for mining, log events captured by the learning environment were mapped to sequences of canonical actions in five primary categories (Kinnebrew, Loretz & Biswas, 2013):

- *READ*: students access a page in the resources;

- *LINK or CONCept Edit*: students edit the causal map, with actions further divided by whether they operate on a causal link ("LINK") or concept ("CONC") and whether the action was an addition ("ADD"), removal ("REM"), or modification ("CHG"), e.g., *LINKREM* or *CONCADD*;

- *QUER*: students use a template to ask Betty a question, and she uses causal reasoning with the current map;

- *EXPL*: students ask Betty to explain her answer to a question;

- *QUIZ*: students have Betty take a quiz.

Results

To determine if our intervention helped students learn the science content and causal reasoning skills, we computed: 1) student pre-to-post learning gain, and 2) students' best causal map scores[3] during the intervention. Table 1 presents these results for each treatment in the intervention. A repeated measures ANOVA performed on the pre- and post-test data revealed a significant effect of time on pre-to-post-test scores ($F = 59.31$, $p < 0.001$, $\eta^2_p = 0.481$), but it failed to reveal a significant effect of treatment ($F = 0.988$, $p > .05$, $\eta^2_p = 0.044$). Similarly, an ANOVA revealed no significant effect of the treatment on the map scores. Clearly all students learned as the result of the intervention and several students produced a significant portion of the correct causal map.

Table 1. Performance [mean (s.d.)] by treatment.

Group	Pre-Test	Post-Test	Gains	Best Map
Con-G	5.07 (2.03)	6.10 (2.64)	1.03 (1.99)	8.87 (8.20)
KC-G	3.85 (2.54)	5.13 (3.37)	1.28 (2.33)	9.55 (6.64)
SE-G	4.41 (1.97)	6.82 (2.33)	2.41 (1.92)	9.53 (7.55)
Full-G	3.88 (1.77)	6.78 (2.76)	2.91 (1.76)	7.25 (6.36)

However, the small sample sizes and the large variations in performance within groups (much more so than across groups) make detailed analysis of the experimental treatments difficult. Therefore, in our application of the analysis framework to data from this study, we focus on analyzing the different learning behaviors corresponding to a given action pattern and comparing the occurrence of these behaviors between students who had high map scores and those who had low map scores, without regard to treatment. The median map score was 7.5, so we consider the students with a map score of 7 or lower as the "*LowMap*" group and the ones with a map score of 8 or higher as the "*HiMap*" group. Below we apply our analysis framework to this data.

The results of the sequence mining on students' action sequences showed that [LINKADD]→[QUIZ] →[LINKREM] was a frequent action pattern across all groups. Initial interpretation of this pattern suggests a behavior in which students use the quiz to check newly added links and then remove incorrect links using the quiz results. Analysis of the effectiveness of the link edits showed that over 88% of the 464 instances of this pattern resulted in an effective link removal (i.e., the link removed was an incorrect one) and 80% of the total involved an ineffective link add. Therefore, we further investigate the 370 instances of the specific pattern [Add Incorrect Link (AIL)] →[Quiz (Q)]→[Remove Incorrect Link (RIL)]. Analysis of the occurrence of this pattern indicates that the *HiMap* group tended to use this pattern primarily in the latter half of their activities, as illustrated in Figure 4. On the other hand, the *LowMap* group tended to use this pattern early on. Since both groups tended to use the individual, component actions similarly over the course of their activities (with somewhat more AIL actions early, more RIL actions late, and somewhat more Q actions late), this indicates a potentially important difference for when, and possibly why, high-performing versus low-performing students employed this behavior. For example, students in the low-performing group resorting to a guess and check behavior early on may have been driven by difficulties in identifying causal relationships in the resources or even a more general disengagement early in the task. Agents in the system can ask questions to better identify the cause(s) of such a behavior and address them with targeted feedback and support. On the other hand, resorting to guess and check behavior late in the task more often indicates that the student believes they have exhaust-

---

[3] The best map score is the highest map score a student achieved at any time during the intervention, calculated as the number of correct casual links minus the number of incorrect causal links.

ed their other approaches to the problem and has fallen back on an inefficient guessing behavior. In such cases, the students may benefit from agent support and feedback on employing more advanced strategies.
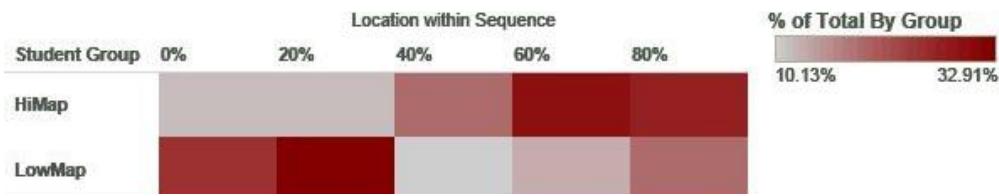
Figure 4. Heat Map illustrating occurrence of AIL → Q → RIL pattern over time.

We built a more detailed understanding of the learner behavior associated with this pattern by considering whether instances of the pattern were coherent and whether the initial [Add Incorrect Link] action was supported. An unsupported action was likely a guess. For this pattern, we employ a strong version of pattern coherence, requiring that the two link edits operated on the same link (i.e., the same link was added and then removed) and that the quiz provided action support for the subsequent link removal. Table 2 shows the number of occurrences of this pattern split by these pattern coherence and initial action support measures.

Table 2. AIL→Q→RIL Behaviors.

Coherent	Support	Behavior Interpretation	Occ.
Yes	Yes	Informed Checking (Eff)	194
Yes	No	Guess and Check (Eff)	52
No	Yes	Informed Checking (Ineff)	71
No	No	Guess and Check (Ineff)	53

Although the coherent, supported version of the pattern dominates (making up over half of the total instances), other versions of the pattern also appear relatively frequently. In general, the initial AIL→Q portion of the pattern illustrates a checking behavior, which is important given that the added link was, in fact, incorrect. Therefore, even when a coherent instance of the full pattern is not found, the AIL→Q sub-pattern still represents a form of checking, albeit an ineffective one because the incorrect link was not subsequently removed. Further, when the initial link add is supported, it suggests that though the edit was incorrect, it was supported by previous information acquisition. Therefore, we consider the supported versions of the pattern or sub-pattern to be an informed checking (SE) behavior, while the unsupported versions represent a less preferable guess-and-check (SE) behavior. Figure 5 illustrates the combination of the information acquisition/application measures with the mined pattern instances to differentiate the corresponding strategies in the learner model.
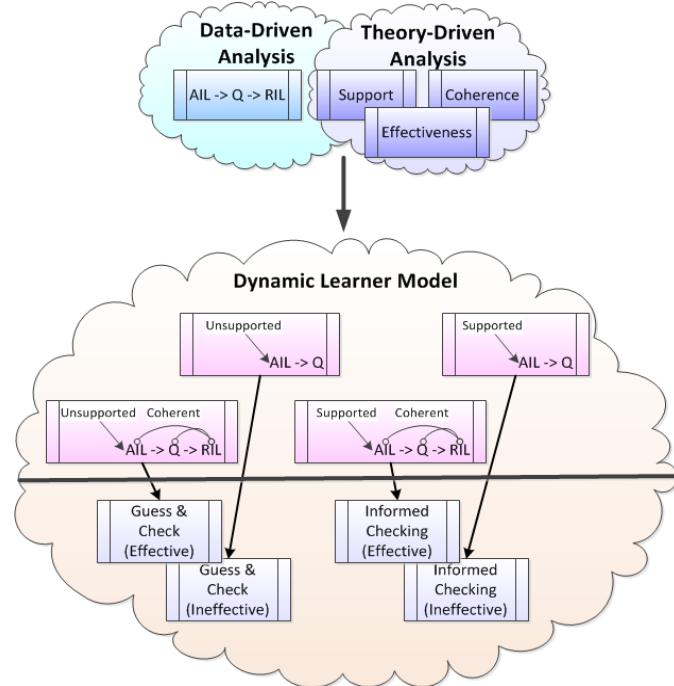
Figure 5. Analysis of AIL→Q→RIL pattern.

There are clear differences in the occurrence of the AIL→Q→RIL pattern between the *HiMap* group, who used the pattern 294 times, and the *LowMap* group, who used it only 74 times. Equally striking are the *HiMap* and *LowMap* groups' relative use of behaviors represented by the different versions of the pattern shown in Table 3. While most of the instances of this pattern correspond to an effective informed checking behavior in the *HiMap* group, the majority of use in the *LowMap* group corresponds to an ineffective guess and check behavior. This supports the hypothesis that one of the major differences between the high- and low-performing students in this study was their ability to employ effective solution evaluation behaviors.

Table 3. *HiMap* versus *LowMap* distinctions.

Behavior	HiMap	LowMap
Informed Checking (Effective)	61%	21%
Guess and Check (Effective)	15%	12%
Informed Checking (Ineffective)	20%	16%
Guess and Check (Ineffective)	5%	51%

Similar analyses of other interesting frequent patterns, such as [Remove Link (RL)]→[Read (R)]→[Add Link (AL)], show that pattern coherence and effectiveness measures allow us to break this pattern down into a number of distinct behaviors. Coherence between the RL, R, and AL actions indicates an attempt by the students to correct their maps. Otherwise, the RL action seems to be independent of the R and AL actions. If the R and AL actions are coherent, the AL is informed, and otherwise the AL action is not informed by the content just read. Overall, only 38% of informed map correction attempts are effective, while the majority (58%) of all informed map additions is effective. This suggests that further support and scaffolding that helps students go back to read the page(s) related to the incorrect link that was just deleted and look for specific information that is related to the correct version of the link just deleted may help students become better at integrating their solution evaluation (i.e., finding errors in their maps) and

knowledge construction activities (i.e., correcting the errors found in the map). In general, we need to provide additional metacognitive strategy-level scaffolds that help students integrate findings from the quiz with targeted information seeking to help them find the correct version of incorrect and missing links in their maps.

Discussion

In this chapter, we have presented a framework for analyzing learning activity data in open-ended learning environments that integrates top-down, theory-driven measures and bottom-up, data-driven pattern discovery. This analysis framework can form the basis for designing richer learner models that characterize students' activities by analyzing their learning behaviors and performance in an integrated fashion. Therefore, the framework advances conventional learner modeling approaches that tend to focus on performance and skills (e.g., Brusilovsky & Millan, 2007; Desmaris & Baker, 2012), and extends learner modeling and analysis beyond step-by-step tutoring systems to more open-ended task analysis, where students are not restricted in their choice of developing problem solutions (e.g., Baker, Corbett & Aleven, 2008; Corbett & Anderson, 1994; Woolf, 2010). We believe that providing students with greater choice allows them to explore a number of alternate solution paths in the solution space, and by self-reflection or guidance from the system develop awareness and discover learning strategies to make better choices and become more effective learners and problem solvers.

Therefore, an important implication of this work that combines discovery of frequent action patterns with action support, pattern coherence, and effectiveness measures in the context of the students' overall activities is the ability to perform much deeper analyses of students' cognitive and metacognitive abilities as they work on their learning and problem solving tasks. This provides opportunities for providing relevant scaffolds that are triggered based on the system's evaluation of the students' recent activities and performance. In past work, we have found that students tend to ignore feedback provided by the system, and very often this is attributed to their inability to understand the feedback, and a lack of understanding of how it will help them overcome the current difficulties that they are facing in the system (Segedy, Biswas & Kinnebrew, 2012). In future work, we will incorporate pattern detectors that are derived from previously identified patterns and the information acquisition/application measures into the Betty's Brain system to directly test the results of this analysis in improving learner scaffolding and engagement with the system.

Recommendations and Future Research

The approach presented in this chapter may provide the basis for developing a metacognitive tutoring framework within the GIFT architecture (Sottilare et al. 2012) to address U.S. Army challenges in computer-based learning, problem solving, and training environments that adapt to the learner's competence and state, while providing "self-development" support for skills that apply across a variety of domains. The metacognitive tutoring framework can adopt a comprehensive approach to developing adaptive modules that support online learning and problem solving by remediating deficiencies in both cognitive skills and metacognitive strategies.

Our suggested approach to metacognitive tutor design within the GIFT framework starts with a specification of a cognitive/metacognitive task model of the training domain using approaches that extend well-known methods like Cognitive Task Analysis (CTA) (Chipman, Schraagen & Shalin, 2000) by expanding the focus to both the cognitive skills and the metacognitive strategies required to achieve proficiency in the chosen domain. The cognitive/metacognitive task model can form the basis for populating the other primary components of the tutoring framework: 1) the learner modeling module; 2) a set of instructional

strategies; and 3) the design of software monitors or sensors to track learner activities and behaviors as they work on the system.

The overall approach to integrating the metacognitive tutoring framework with GIFT by closing the loop in system authoring and analysis is shown in Figure 6. Starting from the authoring tools to support the design and deployment of the cognitive/metacognitive task model, the figure illustrates the sequence for designing and implementing the relevant components of the GIFT tutor. The task model populates the behavior part of the learner model, and the domain model (developed by the training experts) can populate the performance components of the learner model. The task model can also provide a foundation to design and develop the set of instructional strategy templates, which, when populated by the instructional experts, become part of the pedagogical module.
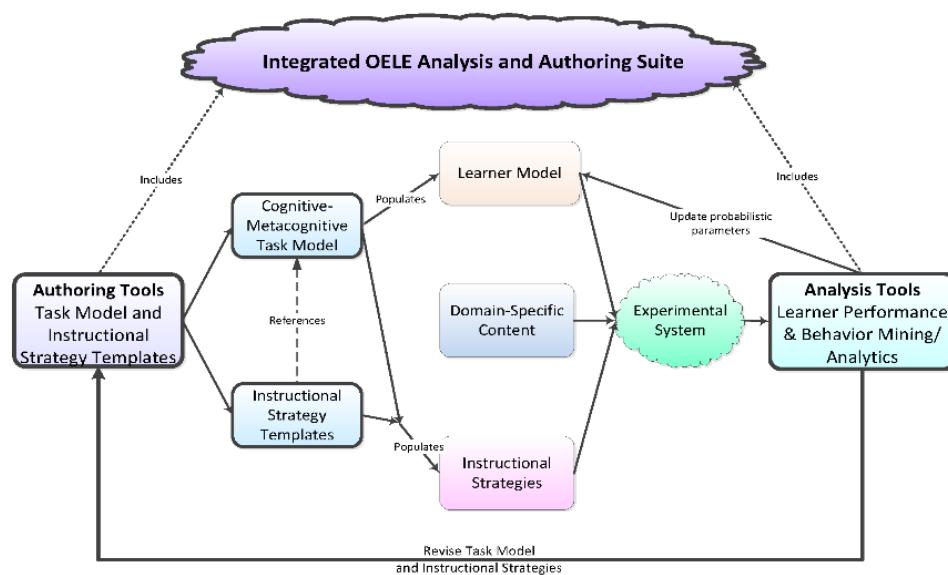
Figure 6. Authoring Toolkit to Support Metacognitive Tutoring in the GIFT Framework.

In future work, we will develop a metacognitive tutor UrbanSim training environment (http://ict.usc.edu/prototypes/urbansim/). UrbanSim is a PC-based virtual training application for practicing the art of mission command in complex counterinsurgency and stabilization operations. In the UrbanSim practice environment, trainees take on the role of an Army battalion commander to plan and execute operations in the context of a difficult fictional training scenario. After developing their intent, identifying their lines of effort and information requirements, and selecting their measures of effectiveness, trainees direct the actions of a battalion as they attempt to maintain stability, fight insurgency, reconstruct civil infrastructure, and prepare for transition. Our first challenge will be to design a relevant cognitive/metacognitive task model in this domain by gaining a deeper understanding of the UrbanSim training scenarios, trainee actions, the decision-making model based on expert-generated policies, and the explanation structures generated to justify the application of policies and evaluate trainee actions. We will use the knowledge gained from this exercise to design three modules: sensor modules, the learner model, and the instructional templates within the GIFT architecture. Populating the instructional templates will provide an initial prototype system that will help us evaluate the effectiveness of our approach in a new OELE environment.

References

Amershi, S. & Conati, C. (2009). Combining unsupervised and supervised classification to build user models for exploratory learning environments. *Journal of Educational Data Mining*, *1*(1):18-71.

Azevedo, R. & Jacobson, M. (2008). Advances in scaffolding learning with hypertext and hypermedia: A summary and critical analysis. *Educational Technology Research and Development*, *56*(1):93-100.

Baker, R.S.J., Corbett, A.T. & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing. In *Intelligent Tutoring Systems*, (pp. 406-415). Springer Berlin.

Bransford, J.D., Brown, A.L., Cocking, R.R. (Eds.). (2000). *How people learn*. Washington, D.C.: National Academy Press.

Brown, A., Bransford, J.D., Ferrara, R.A. & Campione, J.C. (1983). Learning, remembering, and understanding. In P. Mussen, (Ed.) *Handbook of Child Psychology*. Hoboken, NJ: John Wiley.

Brusilovsky, P. & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In *The adaptive web* (pp. 3-53). Springer-Verlag.

Corbett, A.T. & Anderson, J.R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, *4*(4), 253-278.

Desmarais, M.C. & Baker, R.S.J. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, *22*(1-2), 9-38.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American psychologist*, *34*(10), 906.

Hennessey, M.G. (1999). Probing the Dimensions of Metacognition: Implications for Conceptual Change Teaching-Learning. *ERIC*.

Ho, J., Lukov, L. & Chawla, S. (2005). Sequential pattern mining with constraints on large protein databases. In *Proceedings of the 12th International Conference on Management of Data* (COMAD), pp. 89-100.

Kinnebrew, J. S. & Biswas, G. (2012, June). Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. In *Proceedings of the 5th International Conference on Educational Data Mining* (EDM 2012). Chania, Greece.

Kinnebrew, J. S., Loretz, K. M. & Biswas, G. (2013). A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining*, *5*(1):190-219.

Kinnebrew, J. S., Mack, D. L. & Biswas, G. (2013). Mining temporally-interesting learning behavior patterns. In S. D'Mello, R. A. Calvo & A. Olney (Eds.). *6th International Conference on Educational Data Mining*, pp. 252-255.

Lajoie, S. & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In P. Alexander & P. Winne (Eds.), *Handbook of Educational Psychology*, second edition, pp. 803-821. Mahwah, NJ: Erlbaum.

Land, S. (2000). Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development*, *48*(3), 61-78.

Leelawong, K. & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education*, *18*(3), 181-208.

Luckin, R. & du Boulay, B. (1999). Ecolab: The development and evaluation of a Vygotskian design framework. *International Journal of Artificial Intelligence in Education*, *10*(2), 198-220.

Luckin, R. & Hammerton, L. (2002). Getting to know me: Helping learners understand their own learning needs through metacognitive scaffolding. In S. Cerri, G. GouardAlres & F. ParagualCu (Eds.), *Intelligent Tutoring Systems, volume 2363 of Lecture Notes in Computer Science*, pp. 759-771. Springer.

Martinez, M.E. (2006). What is metacognition? *Phi Delta Kappan*, 696-699.

Nesbit, J., Zhou, M., Xu, Y. & Winne, P. (2007). Advancing log analysis of student interactions with cognitive tools. *12th Biennial Conference of the European Association for Research on Learning and Instruction* (EARLI).

Park, O.C. & Lee, J. (2004). Adaptive instructional systems. In D. H. Jonassen (Ed.), *Handbook of research for education communications and technology* (2nd ed., pp. 651–684). Mahwah, NJ: Erlbaum.

Perera, D., Kay, J., Koprinska, I., Yacef, K. & Zaiane, O. Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, 21(6):759-772, 2009.

Puntambekar, S. & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(1):1-12.

Romero, C. & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics*, 40(6):601-618.

Segedy, J., Loretz, K. & Biswas, G. (2013). Model-driven assessment of learners in an open-ended learning environment. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 200-204. New York, NY: ACM.

Segedy, J. R., Biswas, G. & Sulcer, B. (2014). A model-based behavior analysis approach for open-ended environments. *Journal of Educational Technology & Society*, 17 (1), 272–282.

Spires, H., Rowe, J., Mott, B. & Lester, J. (2011). Problem solving and game-based learning: Effects of middle grade students' hypothesis testing strategies on learning outcomes. *Journal of Educational Computing Research*, 44(4):453-472.

Su, J.-M., Tseng, S.-S., Wang, W., Weng, J.-F., Yang, J. & Tsai. W.-N. (2006). Learning portfolio analysis and mining for scorm compliant environment. *Journal of Educational Technology and Society*, 9(1):262-275.

Tang, T. & McCalla, G. (2002). Student modeling for a web-based learning environment: A data mining approach. In *Proceedings of the 18th National Conference on Artificial Intelligence*, pages 967-968. Association for the Advancement of Artificial Intelligence, AAAI Press.

Veenman, M.V. (2012). Metacognition in Science Education: Definitions, Constituents, and Their Intricate Relation with Cognition. In *Metacognition in Science Education*, pp. 21-36. Springer Netherlands.

Winne, P. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and individual differences*, 8(4):327-353.

Winne, P.H. & Hadwin, A.F. (1998). Studying as self-regulated learning. In D.J. Hacker, J. Dunlosky & A. Graesser (Eds.) *Metacognition in educational theory and practice*, Lawrence Erlbaum Associates Publishers, pp 277-304.

Winne, P.H. & Hadwin, A.F. (2008). The weave of motivation and self-regulated learning. In D. Schunk, B. Zimmerman (Eds.) *Motivation and self-regulated learning: Theory, research, and applications* (pp 297-314). NY: Taylor & Francis.

Woolf, B.P. (2010). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. CA: Morgan Kaufmann.

CHAPTER 12 – **Macro and Micro Strategies for Metacognition and Socially Shared Regulation in the Medical Tutoring Domain**

Susanne P. Lajoie and Eric Poitras
McGill University

Abstract

Current models of self-regulation in problem-solving describe efforts to monitor and control several types of processes as an iterative cycle of forethought, performance, and reflection. In this chapter, we outline theoretical constructs that account for how medical students monitor and adaptively control the cognitive, behavioral, motivational, and affective processes that mediate diagnostic reasoning processes and outcomes in the context of clinical reasoning and communication. We draw on several case examples where non-adaptive and adaptive instructional strategies are used to support self-regulation in the context of technology-rich learning environments. To do so, we distinguish between instructional strategies that target processes at the meta-level (planning, orienting, executing, monitoring, evaluating, and elaboration) and object-level (cognition, behavior, motivation, and affect). We further characterize these constructs in terms of levels of granularity (macro vs. micro) and valence (appropriate vs. inappropriate). We elaborate on cases where self-regulation occurs in individual or group settings by explaining intrinsic and extrinsic regulatory processes in the context of physician-patient communication. Finally, we explore the nature of team co-regulation and socially shared regulation in the context of trauma and medical emergency situations.

Introduction

Technology-rich learning environments enhance learning and performance by providing learners with tools that are designed to facilitate the attainment of instructional objectives. The manner in which learners interact with these tools, whether mediated through computer or human tutoring, can be studied to understand how to best assess these interactions and how to intervene when deemed appropriate. In doing so, tutors adaptively respond to the different needs, preferences, and challenges that face individual learners, and select the most suitable type of instruction. In this chapter, we use the term "adaptation" to describe this type of instructional technique, although researchers from the learning sciences community have also used synonymous terms such as "personalization" and "individualization."

Adaptation is at the center of an ITS approach. The success of many ITSs can be attributed to the quality of the CTAs (Anderson & Schunn, 2000) that led to accurate learner and performance models that drove the adaptive instructional strategies (Lajoie & Azevedo, 2006). The outcome of a CTA includes decomposing elements of the domain so that instruction can be designed around these elements and that learners can be assessed when they interact with the instructional material. Our contention is that there is a need to create instructional strategies that are somewhat malleable to account for the multiple solution sequences that people take to solve problems.

The widespread interest in the instructional benefits of adaptation have led to intensive research over several decades (Kay & McCalla, 2012; Azevedo & Aleven, 2013); Some researchers have relied on metaphors, such as the computers as cognitive tools metaphor (Lajoie, 2000; 2005), to guide the design of adaptive capabilities of technology-rich learning environments. For example, they target 1) adaptive
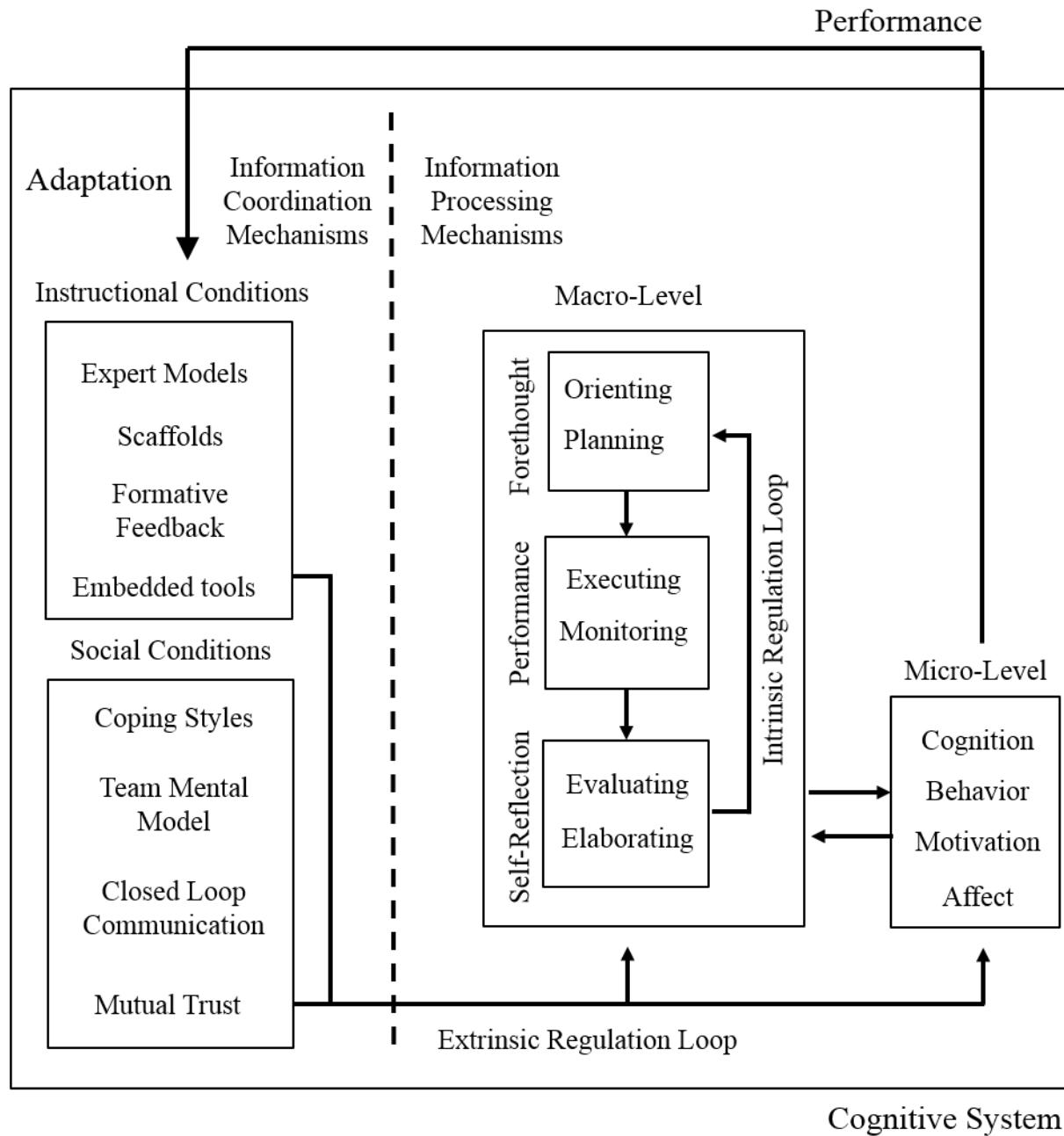
instructional strategies to monitor and react to specific learner characteristics and proficiencies by adapting assistance to these needs; and 2) non-adaptive strategies that rely on the nature of the cognitive tools embedded in the learning environment to scaffold the learner by constraining the problem solving based on relevant actions.

In the case of both ITS and cognitive tools, there is a need to increase the scope of adaptive systems in terms of a more complete examination of learner characteristics, spanning from learner knowledge, cognition, and metacognition to motivation, emotion, and affects. As such, the goal of this chapter is to expand the metaphor of using computers as cognitive tools by demonstrating the use of adaptive and non-adaptive instructional strategies in the context of technology-rich learning environments for medical education. In particular, we focus on the characteristics of regulatory activities, a 21st century skill that is critical for learners to master in the medical domain. In the following sections, we first define and differentiate the characteristics of intrinsic and extrinsic regulatory activities within individual learners (i.e., self-regulation) and groups of learners (i.e., co and socially shared regulation). These constructs are further characterized in terms of levels of granularity (macro vs. micro) and valence (appropriate vs. inappropriate). We then distinguish between instructional strategies that target processes at the meta-level (planning, orienting, executing, monitoring, evaluating, and elaboration) and object-level (cognition, behavior, motivation, and affect). In doing so, we analyze relevant instructional strategies in the context of technology-rich learning environments, drawing on several tasks that include diagnostic problem solving, patient management in trauma and medical emergency situations, and communication of bad news.

Related Research

The breadth and depth of conceptual elements included in contemporary theories of SRL has grown increasingly sophisticated over the past decade (Pintrich, 2004; Schunk, 2005; Winne & Hadwin, 2008; Zimmerman, 2006, 2008; Zimmerman & Schunk, 2011). The focal constructs of SRL theories characterize learners' efforts to monitor and control their own learning (Dinsmore, Alexander & Loughlin, 2008; Lajoie, 2008). As such, SRL is conceptualized as a recursive process that unfolds before, during, and after a learning episode, and is a superordinate construct to metacognitive knowledge and activities. Although theories of SRL share some common assumptions (Pintrich, 2000; Zimmerman, 2001), the choice of constructs reflects inherent differences in the nature of the domain or task (Alexander, Dinsmore, Parkinson & Winters, 2011; Meijer, Veenman & van Hout-Wolters, 2006; Poitras & Lajoie, 2013) as well as relevant environmental and contextual conditions, whether learning occurs within an individual or group (Hadwin & Järvelä, 2011; Järvelä & Hadwin, 2013; Volet, Vauras, Khosa & Iiskala, 2013).

Our research in the regulation of problem solving in the medical domain has led to a synthesis of existing models to account for the relevant domain knowledge (Meijer et al., 2006; Lu & Lajoie, 2008; Lajoie & Lu, 2012; Järvelä & Hadwin, 2013; Volet et al., 2013). In doing so, the choice of constructs is guided by the declarative and procedural knowledge that is inherent to the domain of medical education, and how it mediates efforts to monitor and control performance on several problem-solving tasks, as shown in Figure 1. It is our contention that there are domain-specific SRL skills that must be operationalized in order to promote them with adaptive instruction.

Figure 1. Intrinsic and extrinsic regulation in medical problem-solving.

SRL follows a social cognitive perspective and consists of cognitive, affective, motivational, and behavioral activities that are planned and adapted for the purposes of goal attainment (Zimmerman, 2000; Zimmerman and Campillo, 2003). That being said, one must consider both the intrinsic and extrinsic influences on SRL. From an intrinsic perspective, self-regulated problem solvers engage in cycles of forethought, performance, and reflection (Zimmerman, 2000). *Forethought* refers to the thoughts and beliefs held by novices prior to performance as well as the relevant task conditions that can subsequently affect problem solving, with the *performance* phase involving the steps taken to solve the problem that are monitored and controlled, and *self-reflection* consisting of the novices' judgment and reaction to

performance. The problem-solving process is recursive in that the outcomes of prior steps inform the next ones that are taken to solve the problem. The intrinsic regulatory loops are facilitated by information processing mechanisms that are either basic general skills or domain-specific skills that lead to the transformation of information during problem solving (Winne & Hadwin, 2008). Extrinsic regulation consists of environmental conditions, i.e., the instructional conditions or the social elements that faciliate self-regulation but are not yet internalized in the learners' cognitive system (Gross & Thompson, 2007). We limit the scope of this chapter to the instructional and social conditions that mediate both types of regulatory loops by examining relevant examples from the medical domain. In the following sections, we define the theoretical constructs and processes involved in both the intrinsic and extrinsic regulation loop.

Intrinsic Regulation of Learning in Medical Training

The deployment of regulatory processes can be examined at different levels of granularity and valence as it unfolds during task performance (Azevedo, 2009; Greene & Azevedo, 2010). As such, we first characterize the intrinsic regulation loop in terms of coarse-grained constructs (i.e., the meta-level processes) that are involved in monitoring and adaptively controlling certain aspects of problem solving. These aspects refer to fine-grained constructs (i.e., the object-level processes), whether cognitive, behavioral, motivational, or affective in nature.

As an example, a typical learner will begin to solve a problem such as diagnosing a patient condition by orienting themselves within a problem space and planning the necessary steps to reach a solution, a phase referred to as *forethought*. The learner notices that the patient's heart rate exceeds the normal range and could potentially be caused by a tumor of the adrenal glands. To test this assumption, the plan might entail testing for pheochromocytoma by ordering a lab test to verify serum levels of the catecholamines adrenalin and noradrenalin. At the *performance* phase, self-regulated learners should execute steps to solve the problem and monitor the outcomes. In doing so, the lab test was found to be pertinent, as serum levels were elevated, thereby confirming a diagnosis of pheochromocytoma. The learner then evaluates the progress made in solving the problem by re-adjusting the plausibility of differential diagnoses. In the final *self-reflection* stage, self-regulated learners re-evaluate the state of the problem or elaborate their final solution. Reflections about the problem may lead a learner to order a battery of tests to rule out commonly known alternative diagnoses to pheochromocytoma or they may decide to proceed to begin the relevant treatment plan. Self-reflection processes occur after the performance phase, and, in turn, influences forethought in relation to subsequent steps taken to address the problem. The phases of self-regulation thus involve several types of meta-level processes, namely, orienting, planning, executing, monitoring, evaluating, and elaborating.

Valence is determined by the impact of the intrinsic regulation loop toward task performance, distinguishing between processes that are appropriate or inappropriate given the demands of the situation. The characteristics of the problem space can directly impact how learners should regulate their problem solving, for instance, depending on whether the patient is stable or has deteriorating vital signs. On the one hand, learners who diagnose a stable patient can orient themselves more extensively in the problem space by considering all available information. Self-regulated problem solvers formulate a differential diagnosis that consists of a list of plausible diseases based on a careful consideration of the case history, including symptoms and vital signs, lab tests results, and environmental factors. More time spent in the forethought and self-reflection phase is deemed appropriate since it is conducive to superior performance. On the other hand, learners who must stabilize a deteriorating patient must orient themselves in a different manner, relying on heuristic approaches to address medical emergencies. Instead of considering all factors involved in patient care, the heuristic allows learners to focus on the most life-threatening clinical problems, such as first checking and clearing the airway of a patient, in order to be more efficient in formulating an action plan. As such, the appropriateness of SRL process is determined by the demands of

the problem space. Adaptive SRL skills are based on recognizing the dynamic nature and context of the medical situation and establishing goals based on this awareness.

Extrinsic Regulation of Learning in Medical Training

The extrinsic regulation loop reflects the situation in which learning occurs, as problem solving often involves groups and teams of learners that coordinate and share efforts to monitor and control task performance (Lajoie & Lu, 2012). The extrinsic regulation loop thus captures the influence of social interactions, since learners' plans and goals, standards for monitoring progress, as well as the use of strategies and tactics can be co-constructed with the help of other learners (Hadwin & Oshige, 2011; Hadwin & Järvelä, 2011). Researchers have characterized extrinsic regulation by differentiating between co-regulated learning and socially shared regulation. According to Järvelä and Hadwin (2013), co-regulated learning occurs when a group of learners interact to shape, guide, and constrain learners' regulatory activities. Socially shared regulation refers to a situation where a group of learners collectively shares and constructs goals as well as negotiates standards for judging progress and strategy use in the service of a shared outcome. Both co-regulation and socially shared regulation can be conceived as distinct from collaborative task outcomes or levels of engagement, since the extrinsic regulation loop focuses on meta-level activities, and how these are co-constructed within a group or team of learners.

Researchers have expanded the definition of socially shared regulation by examining social dynamics and relationality of individuals toward each other within a group (Volet et al., 2013). In any group activity, there are social dynamics where an individual must make continuous situational and developmental adjustments to his or her own behavior as the activity changes. One must also consider the relationality of how one interacts toward others in certain contexts and how this will affect the extent of socially shared regulation. The appropriation of self-regulation is not limited to working with more capable group members, as technology can also support group interactions through embedded tools (Hadwin, Oshige, Gress & Winne, 2010). For instance, tools that are designed to support extrinsic regulation might include chat windows, shared digital resources, and artificial pedagogical agents that serve as coaches and facilitators. As such, these social and instructional conditions impact the extrinsic regulation loop by mediating how information is coordinated among group members. The line between SRL, co-regulation and socially shared regulation is somewhat blurred depending on the specific event and situation. We provide some examples in the medical context below.

As an example, learners communicate unfavorable news to simulated patients by regulating their own emotional reactions as well as adaptively responding to the needs and reactions of patients (Lajoie et al., 2012). Information coordination mechanisms involve appropriate efforts on the physician's part to match information giving with the patient's willingness to receive information and how they cope with the situation (Lazarus & Folkman, 1984; Miller, 1995; Carver, Scheier & Weintraub, 1989). The medical curriculum includes opportunities for physicians to learn how to conduct a medical interview with a patient. Part of their training is to identify the patients' coping styles and respond with an appropriate level of information that is adaptive to that patient. Self-regulated problem solvers provide detailed and succinct information about the medical condition to patients who have a monitoring coping style that is problem-focused. However, some patients have a blunting coping style and tend to avoid threat-related information when in distress. In this case, an adaptive response by the learner is to provide the patient with less information, avoid jargon, and rely more on empathetic responses. Some patients have an emotion-focused coping style in response to the stressful situation. In this case, the learner should empathize with the patient rather than provide information about the disease.

Extrinsic regulation is also prevalent in the context of communications among groups and teams of learners in order to ensure successful task performance in medical trauma situations (Cruz-Panesso,

Lajoie & Lachapelle, 2013; Driskell, Goodwin, Salas & O'Shea, 2006; Mathieu, Heffner, Goodwin, Salas & Cannon-Bowers, 2000; Salas, Sims & Burke, 2005). In the early stages of problem solving, the aim of communication is to establish a team mental model, where a knowledge structure is shared across team members, allowing them to share a similar understanding of the task. Knowledge structures that are both accurate and shared across team members allow them to communicate implicitly as well as to make similar inferences and predictions. Information is effectively shared across physicians during task performance through closed-loop communication patterns. In doing so, a message is transmitted by the sender to be interpreted and acknowledged by the receiver. The message itself should contain the proper terminology and be succinct, clear, and audible. Trust among team members develops over repeated practice in solving problems, as learners reflect on task performance, and impacts future communication between team members. Mutual trust is essential since efforts to monitor task performance can be allocated to critical aspects of the task and responsibilities are shared across team members. In the following section, we elaborate further on the role of instructional conditions in facilitating extrinsic and intrinsic regulation, and how instruction can be adapted on the basis of learner performance.

Fostering Self- and Co-Regulation with Technology-Rich Learning Environments

This section describes some technology-rich learning environments that were designed to improve self- and co-regulation in the context of learning in the medical tutoring domain. The *BioWorld* system was developed to allow physicians to practice medical diagnosis and improve their own skills by receiving automated and formative feedback (Lajoie, 2009). The second system, referred to as the *Deteriorating Patient*, simulates a patient with deteriorating vital signs, where the physician must take the appropriate steps to stabilize the patient (Wiseman & Snell, 2008). In the *SimCenter*, physicians receive team training that focuses on combat casualty care, in addition to opportunities for feedback on effective communication and task coordination practices (Cruz-Panesso, 2011; Cruz-Panesso, Lajoie & Lachapelle, 2013; Lajoie, Cruz-Panesso & Lachapelle, in press). *EmpathTools* refers to an online platform that facilitates synchronous discussions between physicians, instructors, and patient actors from different cultures, revolving around the delivery of unfavourable news with the intent of fostering empathy when needed (Lajoie et al., 2012). These learning environments illustrate how one might provide instruction in relation to regulating aspects of learning and task performance in the context of GIFT.

Diagnosing Patients with BioWorld

An important challenge in developing expertise is to provide novices with sufficient opportunities to practice the skills that are necessary to become proficient at a given task. BioWorld is a computer-based learning environment designed to train novices in solving problems in the medical domain (Lajoie, 2009). Novice physicians use BioWorld to diagnose virtual patient cases by identifying relevant symptoms, ordering lab tests, and reasoning about the nature of the underlying disease. Students receive assistance when they request a consult while problem solving. However, the most explicit feedback is provided by BioWorld at the completion of the problem where novices are provided with an explicit representation of each step taken to solve the problem, allowing them to recognize and reflect about where their own solution path differed from the one of an expert.

Physicians rely on the tools embedded in BioWorld to progress through the different phases of regulating problem solving. A typical learner begins to plan their actions by performing a differential diagnosis with the manage hypothesis tool, a dropdown list menu where users select diagnoses, update their confidence, and link relevant information. Learners highlight patient symptoms from the case description to orient themselves to the problem space. The information selected by the learner is stored in the evidence table, a tool designed to support metacognitive monitoring during task performance since it records all the information that was found to be pertinent in solving the case. Experts were found to be more selective

while monitoring their own problem solving, adding less information than the novices to the evidence palette. At the same time, novices selected different information than the experts while solving complex cases, such as pheochromocytoma, for which they might require additional instruction when reaching an impasse (Lajoie et al., 2013). Data-mining techniques have been used to identify these unique patterns. Future work will build on these findings to provide more adaptive feedback at the time of these impasses.

On the basis of the symptom list, the learner updates the confidence meter, a track bar that illustrates the level of confidence toward the main diagnostic opinion, and begins to order lab tests from the patient chart, with the aim of ruling out alternative diagnoses. When the learner reaches an impasse in solving the problem, a library is made available to allow the learner to gather additional information about the diseases and lab tests. The log-file data show that the topic that was consulted by learners in the library can be indicative of misconceptions or impasses that occur while solving particular cases, which can be tracked by the system in order to intervene (Lajoie & Poitras, 2014). For instance, topics such as pheochromocytoma as well as allergies and myocardial infarction are indicative of misconceptions while solving the cases of Amy and Cynthia, respectively. However, information about urinary catecholamine tests and pheochromocytoma are helpful to resolve the Cynthia case, and hyperthyroidism for the Susan Taylor case.

Alternatively, the learner can request a consult as BioWorld can deliver hints in increasing order of specificity. Learners typically ask for consults later in solving a case, particularly when solving more complex cases such as pheochromocytoma. Consult requests typically occur following a lab test result, when learners are 2.1 times more likely to monitor their efforts to solve the problem. Although learners often ask for hints after monitoring their own lack of progress in solving the problem, more proficient novices use this tool to support their own self-reflection, attempting to rule out alternative diagnoses (Lajoie et al., 2012; Lajoie et al., 2013).

Before submitting a final diagnosis, learners can refer to the evidence palette as well as the manage hypothesis tool in order to evaluate their own progress in solving the problem. BioWorld then supports learners in further elaborating the solution by categorizing and prioritizing the evidence items, in addition to justifying the diagnosis by writing a case summary. The final tool is the feedback palette that is designed to support self-reflection, as learners review the evidence items found, and how the items differed from the ones included in the expert solution path to solving the problem. A student report can be obtained with a detailed explanation of the expert solution, including the symptom list, pertinent lab tests, and the differential diagnosis process.

The Deteriorating Patient Activity

As medical students begin to do their clinical work and become responsible for a patients' well-being, they need to understand what they know and what they do not know. One indicator of metacognition in this context is knowing when to ask for help from a senior physician. The Deteriorating Patient (DP) activity (see Wiseman & Snell, 2008 for full description) was created as a simple role-play simulation within a safe classroom environment that provides learners with deliberate practice and feedback as they manage a simulated patient who deteriorates rapidly if the student does not use appropriate medical procedures. This is a human role-play activity that is scripted by the instructor. The objective of the DP activity is to help students learn how to gain control of an unstable patient by applying the appropriate medical algorithm (the ABCDEFG algorithm) to stabilize the patient by checking airways, breathing, circulation, drugs, endocrine/electrolyte, fever, and general, providing the correct medication, conducting appropriate diagnostic tests, etc. The clinical teacher acts as the "deteriorating patient" and responds to the actions of the student physician by recovering or further deteriorating. The instructor provides guidance and hints to scaffold the learner to ensure that the patient survives.

The DP activity can be run as a 1:1 tutoring activity or it can be run with small groups who co-construct their answers in trying to manage the patient. The teachers' role is identical in both situations; however, he is tutoring a group instead of an individual. In the group situation, the students must negotiate the best way to respond to the patient. Lajoie and Lu (2012) examined groups of learners who solved the DP with or without the assistance of technology. Both conditions used a whiteboard to post their medical argument to support their problem list of what the patient might have wrong. However, the technology group had an interactive whiteboard that could be used to share the arguments between the teacher-student as well as between students within and across various teams. The goal of their study was to see whether co-regulation was supported in both collaborative learning situations. The assumption was that groups of individuals are multiple self-regulating agents that socially regulate each other's learning (Volet, Summers & Thurman, 2009). To see whether this assumption was true, group discourse and whiteboard annotations were analyzed to document the presence of SRL and co-regulation, as well as effective patient management. The group discourse was coded to determine the presence of micro-processes of SRL (orienting, planning, monitoring, elaborating, executing, and evaluating).

Similar overall levels of metacognitive activity were found in both conditions but the pattern and timing of metacognitive categories varied, as did patient management. In particular, the technology groups engaged in more planning and orienting at the outset of the problem and this early engagement led to shared understandings and effective patient management in latter sessions. The non-technology groups did not reach a shared understanding as quickly and consequently did not do as well in managing the patient case. The added value of the technology tools was the early facilitation of co-regulation. Specifically, groups oriented to the situation quickly and were able to make decisions more rapidly in this high stress situation than those not supported with technology.

Technology helped facilitate common ground early on in the emergency situation, which is crucial to foster both communication and appropriate actions for patient management. Although both group conditions had a white board to document their plans for solving the case, the technology condition had the benefit of a shared whiteboard on their laptops. The whiteboard had pull-down menus that could be used to annotate different parts of the problem list. The technology used these menus to annotate their problem list, which helped orient each other to the important plans for solving the case. Examining both the types and timing of metacognitive activities can help us identify the points where co-regulation occur and where scaffolding is needed to improve the learning situation.

SimCenter

Medical simulation centers are an excellent context for studying medical teams. Paris, Salas, and Cannon-Bowers (2000) describe a team as two or more people who interact dynamically, interdependently and adaptively toward a common goal, who have been assigned specific roles and functions to perform (Paris et al., 2000). Medical teams work in rapidly evolving and ambiguous situations and they often work with team members that they have not worked with before. Shared social regulation is a necessity in establishing an effective team that can perform under intense time pressure where the patients' well-being is at stake. The medical team works toward a collective purpose by co-constructing what needs to be done to save the patient.

We studied medical military trauma teams who were being deployed overseas after training on several medical emergency scenarios in the SimCenter. These teams had never worked together before and thus their level of familiarity with each other was slight. The goal of this research was to train teams quickly and work effectively and to examine what factors distinguished the least and most effective teams (Cruz-Panesso, 2011; Lajoie, Cruz-Panesso & Lachapelle, in press; Cruz-Panesso, Lachapelle & Lajoie, 2011). Team-based simulations consist of a patient mannequin that mimics patient responses to the team's

actions using high fidelity medical equipment. The role of the team leader is to provide instructions and assign goals based on the information that is received from other team members. Cruz-Panesso (2011) found that teams that implement coordination strategies throughout their performance are more effective for solving simulation team-based scenarios. Coordination strategies refer to managing the interdependencies between activities to achieve a common goal (Malone & Crowston, 1990). The management of dependencies (e.g., roles, tasks, members) can be achieved through implicit and explicit coordination mechanisms (Entin & Sefarty, 1999; Espinosa, Lerch & Kraut 2004; Manser, Harrison, Gaba & Howard, 2009) that allow team members to anticipate others actions and articulate plans and actions to others. These mechanisms lead to shared mental models, closed-loop communication, and the development of mutual trust (Salas et al., 2005). When breakdowns occur in these coordination mechanisms, errors can occur. Thus, team training should identify where coordination fails and why, and provide scaffolding to help teams co-regulate and share regulation appropriately. Although there is an assumed medical hierarchy, where the leader is meant to lead, trust may break down in the group when there are communication failures and the patient is deteriorating. At this point, other role-players may intervene to try to save the patient.

EmpathTools

Medical communication requires a different kind of regulation. It involves emotional regulation and socially shared regulation. For example, the physician or medical student must learn to regulate their own emotions (intrinsic emotions) as they communicate bad news to a patient and they must regulate the emotions of their patients (extrinsic emotions) as well. Communication requires a speaker and a listener where information is conveyed and heard. If the listener is not prepared to hear the message, then the message will not be processed. Using an online video-conference platform, Adobe Connect, we studied the emotional regulation of medical students as they communicate bad news to standardized patients (actors who play the role of a patient). The environment was created to provide students with practice opportunities for communication as well as learning opportunities using a problem-based learning (PBL) intervention that provided video cases of physicians communicating bad news to patients (Lajoie et al., 2013; Hmelo-Silver et al., 2013). These video cases served as the context for discussion about the parameters of effective communication.

Communicating bad news requires that students monitor and control cognitive and affective activities. The cognitive skills pertain to knowledge and understanding of the disease. The affective skills require monitoring their own and their patients' emotions as they give bad news. We examined physician behaviors and patient coping styles to see whether the type of behavior a physician displayed matched appropriately to the type of coping style the patient exhibited. Learning outcomes were explicitly defined as the medical students' ability to monitor patients' coping styles and match their communication behaviors according to the patient's emotional coping strategies.

A case study approach was used given the nature of the data (discourse and text) and small focus groups. Data were analyzed using mixed methods. State-trait analyses revealed the connection between student knowledge and empathy, and the changes in this relationship due to practice within the EmpathTools environment. In this context, we focused on how well the medical students were able to identify the patient's particular coping style as a determinant in choosing the correct information giving responses. The frequencies of the matches between patient coping style and physician communication behavior were identified for each participant pre and post PBL interview. Due to our small sample size, no significant results were obtained on Chi square contingency tests; we did, however, see a decrease in the number of mismatches students obtained in their post-test interviews. EmpathTools was an effective practice environment, helping medical students become more aware of patient's informational needs and emotions and allowing them to tailor their interactions with patients by responding more appropriately based on

their ability to identify patient coping styles. This small case study with human tutors, physicians, and patients has helped us identify the key constructs that lead to more appropriate means of communicating bad news to patients. Eventually, this type of tool could be translated into an ITS if natural language processors were paired with pedagogical agents that could provide the trainees with feedback on their ability to monitor patient response appropriately. What we have found so far is that medical students need to realize that "knowing" what the disease is and transmitting their knowledge of the disease to a patient may not be the appropriate communication strategy for a particular patient.

Discussion

This chapter has demonstrated the manner in which SRL can be explored in ill-structured problem contexts. In particular, we examine SRL in the medical domain using computer-based learning environments. More importantly, we have discussed the macro- and mico-levels of SRL in specific medical contexts. Some of these contexts revealed the importance of individual learning trajectories and the manner in which adaptive instructional strategies could be beneficial. In other contexts, we described the manner in which co-regulation could be assessed and scaffolded using technology in group and team learning situations. Finally, examples of socially shared regulation were discussed with an emphasis on intrinsic versus extrinsic emotion regulation and the development of shared mental models. In each of these examples, we have presented evidence of SRL contingent on the granularity of the problem space and the context in which it was defined. We summarize the instructional strategies that lead to instructional design guidelines for each medical example in Table 1. These strategies vary based on the nature and purpose of the environment and whether it required self-, co-, or socially shared regulation strategies. In the following section, we outline several recommendations for GIFT and identify challenges in delivering instruction to foster self- and co-regulated learning in the medical domain.

Table 1. Instructional strategies that led to design guidelines for each medical example.

	Expert Models	Scaffolds	Formative Feedback	Embedded Tools
BioWorld	Novice-expert overlay of solution paths is based on all actions taken to solve a case.	Hints delivered upon request.	Summary report of novice-expert differences in solution paths and case summaries	Hypothesis manager supports planning and orienting and provides links to the evidence palette that is both a memory and metacognitive tool for supporting plans and their evaluation dynamically. A confidence meter is used to indicate strength of belief in diagnosis. Patient history and vital signs are available in the case and students indicate what they see as relevant. We monitor their awareness and accuracy. Online medical library: do students use it for knowledge acquisition? Patient Chart: do they order, monitor and evaluate diagnostic lab tests? Written case summary where student elaborates patient diagnosis by explaining for the next physician.
Deteriorating Patient	Human medical tutor (subject matter expert [SME]) scripted the case and responded as the patient would based on student's actions. Medical algorithm ABC (Airway-Breathing-Circulation) used to recognize urgent situations and prioritize their diagnostic and therapeutic approaches. Algorithm served as expert model.	Tutor assistance provided: 1. if help requested 2. by improving or deteriorating patient's vital signs based on student actions In group condition feedback by peers on plans, actions, monitoring, evaluating, and elaborating on each other's ideas in order to make appropriate decisions with the patient. Tutor intervened accordingly.	Debriefing of every case by human tutor going over the final problem list created to solve the case.	Structured whiteboard with pull-down menus for annotating the problem list.

Table 1. Instructional strategies that led to design guidelines for each medical example (continued).

	Expert Models	Scaffolds	Formative Feedback	Embedded Tools
Sim Center	Content and pedagogical SMEs created the surgical scenarios (e.g., patient stepped on an explosive device) and programmed mannequins and high fidelity equipment to respond to the team's actions based on case.	Teams scaffolded each other through explicit or implicit communication or actions on the mannequin based on a shared understanding of what was needed in the context of the surgery Instructors altered vital signs of the mannequin to see how team would respond	Debrief by observing instructors on team and individual performance. Cognitive and behavioral skills as well as trust issues and communication were discussed in terms of which interactions were appropriate or inappropriate and strategies provided for how it could have been done differently.	Vital signs of patient would respond. IV and machines that patient was attached to would respond to actions. Blood and bleeding based on team actions.
EmpathTools	Medical algorithm for how to conduct a patient interview: SPIKES (setting, perception, invitation, knowledge, empathy, summary and strategies) (Buckman, 2005).	Problem based learning approach to facilitating learning in small groups. Medical tutor and group members scaffold each other.	Debrief by medical tutors.	Practice with standardized patients (Actors who follow scripts based on physician communications). Video cases used to situate PBL about how to communicate bad news. Interface allowed students to see and interact with each other and with tutor or patient based on the activity. Reflection activities before and after to reflect on performance.

Recommendations and Future Research

GIFT has a lot to offer researchers in that its extensive authoring capability can lead to the development of new tutoring systems that can serve as experimental test beds for improving learning and performance. The framework is comprehensive in that it considers the learner's cognitions and behaviors, traits, and preferences through multiple forms of data, be they log file data of actions taken within a domain context, physiological or behavioral sensors indicating emotional engagement, or self-report measures in the form of surveys providing for accurate evaluation of the learner's states (e.g., engagement level, confusion, frustration). The goal of this multifaceted data collection is to provide a more robust learner model that will be used by the pedagogical module to administer the best pedagogical strategies for scaffolding the learner. The framework also takes into consideration one-to-one (individual) and one-to-many (collective or team) training experiences.

It would be ideal to work with GIFT researchers on some of the issues that slow down the process of modeling in the medical domain. Part of the inefficiency is that medical diagnosis is an ill-structured task. In a study of medical experts using BioWorld, we found that experts agreed on the medical diagnoses for each patient case; however, they took different paths to the solution and did not always prioritize the importance of specific criteria in the same manner (Gauthier & Lajoie, 2013), making it difficult to develop ideal performance models. There was a significant overlap in the actions taken but not everyone agreed on which actions were most important in their decision making. In this situation, adaptive instructional strategies need to be activated based on a participants' chosen path, delivered through either human or computer tutors. In other words, we need to tutor students throughout their problem solving, helping them self-regulate the important processes needed to reach a correct solution. Having a model of appropriate plans, actions, behaviors, and the strategies to help learners acquire the information needed to reach a solution is important regardless of the path taken to solution.

Recently, we have been using data-mining techniques to identify the most frequent solution patterns from the logged user interactions as a precursor to providing appropriate scaffolding (Lajoie et al., 2013). This technique has been successful in discovering unanticipated learner behaviors, and helps us map common impasses or misconceptions that occur while solving the problem. Our first design recommendation for ill-structured domains is to ensure that solution sequences are classified so that appropriate scaffolding can occur when deemed necessary. This is a necessary step as understanding the needs of different learners is crucial to selecting the most suitable instruction.

Appropriate scaffolding can also be delivered based on time factors, as shown with the DP activity, where patient vital signs slowly deteriorate or improve on the basis of learner interactions. The GIFT approach to developing ITSs could allow for a broad range of assessment and instructional methods customized through varying degrees of human involvement before, during, and after a training session. Scaffolding would need to be provided based on the urgency of the situation. Put simply, even if the ultimate goal is to diagnose the source of a patients' problem, there may be critical issues that take priority over such detective work. For instance, if the patient is having trouble breathing, you must address the breathing issue rather than trying to take a diagnostic test to determine the cause of the patients' problem. The type of scaffold you would provide to a student in this "deterioration" would need to be immediate or the patient would die. A different type of scaffold could occur after the case that would focus on the cognitive, metacognitive, affective, and behavioral issues that should have been considered for that case.

As mentioned above, the first design recommendation pertains to one-to-one tutoring strategies in ill-defined problem solving situations. The one-to-many (collective or team) training framework is powerful but considerations need to be given to the evolving theories of co-regulation and socially shared regula-

tion. In particular, different assessment models need to be derived for group- and team-based contexts depending on the nature of the task and requirements of individuals in the group or team. We found that technology can be used as a means to facilitate assessment in such a manner as to support human tutors in adapting their instruction to the specific needs of different learners. For instance, video playbacks of one's own or another's performance of a task was a useful tool to provide such feedback, as instructors were able to elicit and respond to group members' discussions. Based on appropriate analyses of human dialogue, GIFT could go beyond human tutoring by incorporating its powerful natural language capabilities combining appropriate computer-generated scaffolding as a response to human dialogue.

The chat tools in the EmpathTools project were also helpful in tutoring the instructors by monitoring the flow of discussion and selecting the most suitable conversational prompt. ITSs that rely on open-ended learner models, where performance indicators are made explicit and available to the learner, may prove to be a useful feature in facilitating how instructors intervene during a training session with group or team members (see Bull & Kay, 2007). In order to better improve the assessment capabilities of this approach, we recommend that GIFT allows instructors to select the type of data and sample the most pertinent moments to be shown in an open-ended learner model.

A specific design feature for team-based learning would be the ability to identify, model, and scaffold individual roles that team members' play. For example, in our work with trauma teams, we noted that when individuals in the leader role failed to lead, other team members would either be lost in their roles or try to take the lead in patient management (Cruz-Panesso, 2011; Cruz-Panesso et al., 2012). The team demonstrates shared regulation in trying to manage the patient but if one member fails in their role, compensatory strategies by other players take effect. Can we model this type of team learning with GIFT and if so how would we account for group dynamics, co- and shared regulation? Could GIFT be used with "human others" or would we need "pedagogical agents" to serve as individual group member tutors depending on the learning situations?

Finally, the last and most sensitive issue for expanding learner models is the issue of emotional regulation and the consequent adaptive instructional strategies for managing emotions. The question for designers and researchers is what is the purpose for regulating emotions? Researchers are beginning to identify the complex relationship between learning and affect in an attempt to adapt the learning environment to promote engagement in the learning process (Graesser, Hmelo, Calvo, Azevedo, Woolf, Lester, Johnson, etc.). However, we need to consider both intrinsic and extrinsic emotional regulation when dealing with sensitive communication issues, be it in classrooms or real-world contexts. For instance, physicians need to regulate their emotions as well as their patients in order to give information that is properly processed and received by patients. These types of communication patterns could be examined more broadly, using the GIFT framework to expand the notion of emotional regulation in tutoring systems. To summarize, we need to know the purpose for assessing emotion, and then we can determine what actions to take when specific emotions are detected.

In closing, we provided examples of adaptive and non-adaptive instructional strategies used to support self-regulation in the context of technology-rich learning environments in the medical domain. We illustrated the manner in which both macro- and micro-level processes can be detected and used to generate adaptive instructional strategies. Examples from both individual and group settings were provided to demonstrate the robust nature of SRL, co- and shared regulation. Finally, we addressed the nature of intrinsic and extrinsic emotional regulation. Design recommendations have been provided that could extend the GIFT framework to the ill-structured problem-solving domain.

References

Alexander, P. A., Dinsmore, D. L., Parkinson, M. M. & Winters, F. I. (2011). Self-regulated learning in academic domains. In B. Zimmerman & D. Schunk (Eds.), *Handbook of self-regulation of learning and performance.* New York: Routledge.

Anderson, J. R. & Schunn, C. D. (2000). Implications of the ACT-R Learning Theory: No magic bullets. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 5). Mahwah, NJ: Erlbaum.

Azevedo, R. (2009). Theoretical, methodological, and analytical challenges in the research on metacognition and self-regulation: A commentary. *Metacognition and Learning, 4*(1), 87-95.

Azevedo, R. & Aleven, V. (2013). *International Handbook of Metacognition and Learning Technologies.* New York, NY: Springer.

Buckman, R. A. (2005). Breaking bad news: the s-p-i-k-e-s strategy. *Community Oncology, 2*(2), 138-142.

Bull, S. Kay J. (2007). Student models that invite the learner in: The SMILI Open Learner Modelling Framework. *International Journal of Artificial Intelligence in Education 17*(2), 89-120.

Carver, C., Scheier, M. & Weintraub, J. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology, 56*(2), 267-283.

Cruz-Panesso, I. (2011). *Medical Team Effectiveness In a Simulated-Based Learning Environment.* (Unpublished master's thesis). McGill University, Montreal, Canada.

Cruz-Panesso, I., Lachapelle, K. & Lajoie, S. (2011, November). *The role of leadership in establishing mutual trust and team coordination.* Poster presented at the Royal College's 2011 Simulation Summit. Montreal.

Cruz-Panesso, I., Lajoie, S. P. & Lachapelle, K. (2013). *Team cognitive task analysis: The hidden layers of team coordination.* Royal College's Simulation Summit, Promoting Professional Proficiencies, Vancouver.

Dinsmore, D., Alexander, P. & Loughlin, S. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*(4), 391-409.

Driskell, J. E., Goodwin, G. F., Salas, E. & O'Shea, P. G. (2006). What makes a good team player? Personality and team effectiveness. *Group Dynamics: Theory, Research, and Practice, 10*(4), 249-271.

Entin, E. E. & Serfaty, D. (1999). Adaptive team coordination. *Human Factors, 41*, 312-325.

Espinosa, A., Lerch, F. J. & Kraut, R. (2004). Explicit vs. Implicit Coordination Mechanisms and Task Dependencies: One Size Does Not Fit All. In E. Salas & S. M. Fiore (Eds.), *Team Cognition: Process and Performance at the Inter- and Intra-individual Level* (pp. 107-129). Washington DC: American Psychological Association.

Gauthier, G. & Lajoie, S. P. (2013). Do expert clinical teachers have a shared understanding of what constitutes a competent reasoning performance in case-based teaching? *Instructional Science,* 1-16. [Online First]

Greene, J. A. & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist, 45*(4), 203-209.

Gross, J. J. & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 3-26). New York, NY: Guilford Press.

Hadwin, A. F. & Oshige, M. (2011). Self-regulation, co-regulation, and socially-shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*(2), 240-264.

Hadwin, A. F. & Järvelä, S. (2011). Introduction to a special issue on social aspects of self-regulated learning: Where social and self meet in the strategic regulation of learning. *Teachers College Record, 113*(2), 235-239.

Hadwin, A. F., Oshige, M., Gress, C. L. Z. & Winne, P. H. (2010). Innovative ways for using gStudy to orchestrate and research social aspects of self-regulated learning. *Computers in Human Behavior, 26,* 794-805.

Hmelo-Silver, C., Lajoie, S. P., Khurana, C., Lu, J., Wiseman, J., Chan, L.K. & Cruz-Panesso, I. (2013, August). *Supporting international Problem-based learning using CSCL tools.* Presented at the European Association for Research in Learning and Instruction. Munich.

Järvelä, S. & Hadwin, A. (2013). New Frontiers: Regulating learning ini CSCL. *Educational Psychologist, 48*(1), 25-39.

Kay, J. & McCalla, G. (2012). Coming of age: Celebrating a quarter century of user modeling and personalization: Guest editors' introduction. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research, 22,* 1-7.

Lajoie, S. P. & Lu, J. (2012). Supporting collaboration with technology: Does shared cognition lead to co-regulation in medicine. *Metacognition & Learning, 7,* 45-62.

Lajoie, S. P., Cruz-Panesso, I., Poitras, E., Kazemitabar, M., Wiseman, J., Chan, L. K. & Hmelo-Silver, C. (2012). Can Technology Foster Emotional Regulation in Medical Students? An International Case Study Approach. In M. Cantoia, B. Colombo, A. Gaggioli & B. Girani De Marco (Eds.)*, Proceedings of the 5th Biennial Meeting of the EARLI Special Interest Group 16 METACognition* (p. 148). Milano, Italy.

Lajoie, S. P., Naismith, L. & Poitras, E. (2012). Supporting Metacognitive Processes in Medical Reasoning Using Technology. In M. Cantoia, B. Colombo, A. Gaggioli & B. Girani De Marco (Eds.)*, Proceedings of the 5th Biennial Meeting of the EARLI Special Interest Group 16 METACognition* (pp. 47-48). Milano, Italy.

Lajoie, S., Naismith, L., Poitras, E., Hong, Y., Panesso-Cruz, I., Ranelluci, J. & Wiseman, J. (2013). Technology rich tools to support self-regulated learning and performance in medicine. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies*. Amsterdam, The Netherlands: Springer.

Lajoie, S. P., Poitras, E., Naismith, L., Gauthier, G., Summerside, C., Kazemitabar, M., Tressel, T., Lee, L. & Wiseman, J. (2013). Modeling Domain-Specific Self-Regulatory Activities in Clinical Reasoning. In K. Yacef et al. (Eds.), *International Artificial Intelligence and Education Proceedings* (pp. 632-635). Amsterdam: IOS Press.

Lajoie, S., Hmelo-Silver, C., Wiseman, J., Chan, L., Lu, J., Khurana, C., Cruz-Panesso, I., Poitras, E. & Kazemitabar, M. (submitted). Using online digital tools and video to support international problem-based learning. *The Interdisciplinary Journal of Problem-Based Learning.*

Lajoie, S. P. (2009). Developing professional expertise with a cognitive apprenticeship model: Examples from Avionics and Medicine. In K. A. Ericsson (Ed.). *Development of Professional Expertise: Toward Measurement of Expert Performance and Design of Optimal Learning Environments* (pp. 61-83). Cambridge University Press.

Lajoie, S. & Poitras, E. (2014). *Assessment and Instruction of Self- and Co-Regulation of Medical Diagnostic Processes in Technology-Rich Learning Environments*. Presented at the International Conference of the Learning Sciences, Boulder, CO, USA.

Lajoie, S. P., Cruz-Panesso, I., Lachapelle, K. (in press). Learning in the health sector with simulated systems. In M. Spector (Ed.), *Encyclopedia of Educational Technology*.

Lajoie, S. (2005). Extending the scaffolding metaphor. *Instructional Science, 33*(5-6), 541-557.

Lajoie, S. (Ed.) (2000). *Computers as cognitive tools (vol. 2): No more walls* (pp. 1-430). Mahwah, NJ: Lawrence Erlbaum.

Lajoie, S. P. (2008). Metacognition, self regulation and self-regulated learning: A rose by any other name? *Educational Psychology Review, 20,* 469-475.

Lajoie, S.P. & Azevedo, R. (2006). Teaching and learning in technology-rich learning environments. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (2nd ed.) (pp. 803-821). Mahwah, NJ: Erlbaum.

Lazarus, R.S. & Folkman, S. (1984). *Stress, Appraisal and Coping.* New York: Springer.

Lu, J. & Lajoie, S. P. (2008). Supporting medical decision making with argumentation tools. *Contemporary Educational Psychology, 33*(3), 425-442.

Malone, T. W. & Crowston K. (1990). What is coordination theory and how can it help design cooperative work systems?. (Tatar, Deborah, Ed.).*Proceedings of the Third Conference on Computer-supported Cooperative Work (CSCW).* 357–370.

Manser, T., Harrison, T. K., Gaba, D. M. & Howard, S. K. (2009). Coordination patterns related to high clinical performance in a simulated anesthetic crisis. *Anesthesia and Analgesia, 108*(5), 1606-1615.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273-283.

Meijer, J., Veenman, M. V. J., van Hout-Wolters, B. H. A. M. (2006). Metacognitive activities in text-studying and problem-solving: Development of a taxonomy. *Educational Research and Education, 12*(3), 209-237.

Miller, S. M. (1995), Monitoring versus blunting styles of coping with cancer influence the information patients want and need about their disease. Implications for cancer screening and management. *Cancer, 76,* 167–177.

Paris, C. R., Salas, E. & Cannon-Bowers, J. A. (2000). Teamwork in multi-person systems: A review and analysis. *Ergonomics, 43*(8), 1052-1075.

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientations in learning and achievement. *Journal of Educational Psychology, 92,* 544-555.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-407.

Poitras, E. G. & Lajoie, S. P. (2013). A domain-specific account of self-regulated learning : The cognitive and metacognitive activities involved in learning through historical inquiry. *Metacognition & Learning, 8*(3), 213-234.

Salas, E., Sims, D. E. & Burke, C. S. (2005). Is there a "Big Five" in teamwork? *Small Group Research, 36*(5), 555-599.

Schunk, D. H. (2005). Self-regulated learning: The educational legacy of Paul R. Pintrich. *Educational Psychologist, 40,* 85-94.

Volet, S. E., Summers, M. & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction, 19*(2), 128-143.

Volet, S., Vauras, M., Khosa, D. & Iiskala, T. (2013). Metacognitive regulation in collaborative learning : Conceptual developments and methodological contextualizations. In S. Volet & M. Vauras (Eds.), *Interpersonal Regulation of Learning and Motivation: Methodological Advances* (pp. 67-101). New York, NY: Routledge.

Winne, P. H. & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297-314). Mahwah, NJ: Lawrence Erlbaum Associates.

Wiseman, J. & Snell, L. (2008). The Deteriorating Patient: A realistic but "low-tech" simulation of emergency decision-making. *The Clinical Teacher, 5*(2), 93-97.

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology, 25*(1), 82-91.

Zimmerman, B. J. (2001). Achieving academic excellence: A self-regulatory perspective. In M. Ferrari (Ed.), *The pursuit of excellence through education* (pp. 85-110). Mahwah, NJ: Erlbaum.

Zimmerman, B. J. (2006). Development and adaptation of expertise: The role of self-regulatory processes and beliefs. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 705-722). New York, NY: Cambridge University Press.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Zimmerman, B. J. & Campillo, M. (2003). Motivating self-regulated problem solvers. In J. E. Davidson & R. Sternberg (Eds.), *The nature of problem solving* (pp. 233-262). New York: Cambridge University Press.

B. Zimmerman & D. Schunk, (Eds.). (2011). *Handbook of self-regulation of learning and performance*. New York: Routledge.

CHAPTER 13 – Tutoring Self- and Co- Regulation with Intelligent Tutoring Systems to Help Students Acquire Better Learning Skills

Ido Roll[1], Eliane Stampfer Wiese[2], Yanjin Long[2], Vincent Aleven[2], Kenneth R. Koedinger[2]
[1] University of British Columbia, [2] Carnegie Mellon University

Introduction

A number of studies have found that students who better regulate their learning also achieve better learning outcomes within ITSs (cf., Aleven, McLaren, Roll & Koedinger, 2006; Lester, Mott, Robison, Rowe & Shores, 2013; Mathews & Mitrovic, 2008). These correlational results suggest a causal relationship, according to which supporting students' SRL improves their domain-level learning. Thus, helping students regulate their learning has become an increasing focus within ITSs (for overviews, see Azevedo & Aleven, 2013; Goldberg, chapter 10 in this book; Koedinger, Aleven, Roll & Baker, 2009; Roll, Aleven, McLaren & Koedinger, 2007). Indeed, supporting students' SRL has been found to improve domain learning across topics, activity types, and forms of scaffolding. To name a few examples, supporting SRL using feedback and prompts improves learning in a hypermedia environment (Azevedo et al., 2012); prompting students to self-explain improves their learning in problem-solving environments (Aleven & Koedinger, 2002; Conati & VanLehn, 2000; Hausmann & VanLehn, 2007); and offering students tools with which to organize their exploration supports learning in inquiry environments (de Jong, 2006; van Joolingen, 1998). SRL support plays an especially important role in supporting learning in complex activities. Environments that offer exploratory, open-ended activities offer many benefits for learning, as students engage in authentic problem solving (Roll, 2010; Tobias & Duffy, 2009). However, research has shown that learners are in need of support in these environments (Sweller, Kirschner & Clark, 2006). Offering SRL scaffolding has the potential to support the learning process without short-circuiting critical elements of constructivist instruction.

Here we focus on ITSs that help students acquire better SRL skills. Thus, rather than focusing on learning at the domain level, we focus on the potential of ITSs to help students become better life-long learners. We use the term SRL to refer to the collection of strategies and behaviors that students apply to progress within a learning environment (Dinsmore, 2008; Lajoie, 2008; Winne, 2001). We constrain the scope of this chapter by focusing mainly on cognitive and metacognitive aspects of SRL, acknowledging that motivational and attitudinal aspects are no less important (Pintrich, 1999; Zimmerman, 2010). Investigating SRL processes within ITSs gives us a unique lens with which to evaluate metacognitive and SRL skills, as we study the manifestation of these skills in students' actions. For example, the timing, context, and fashion in which students ask for help can be used to infer their help-seeking skills and their metacognitive knowledge of their abilities. The capacity of ITSs to evaluate SRL using behavioral measures, rather than self-reports, is a strength of the field, and allows us to evaluate SRL at a much finer grain-size (Winne, 2006).

We discuss the tutoring of SRL in ITSs by focusing on the form, objectives, and role of SRL scaffolding. With respect to *form,* we identify several types of scaffolds for SRL. These scaffolds give different levels of agency to the student, that is, offer different levels of autonomy and balance of control in the learning process. As for the *objectives*, while scaffolding SRL can improve domain learning, a more ambitious goal is to help students acquire better SRL skills and attitudes that they can transfer to new learning situations. We identify dimensions of transfer of SRL skills within ITSs and evaluate the success of ITSs in achieving this transfer. Regarding the *role* of SRL support, it is important that students practice

regulating their learning processes. Thus, we propose to view learning in ITSs as an emerging outcome of negotiations and interactions between learners and the system. We discuss this perspective in terms of co-regulation and investigate its implications on the design of SRL scaffolding. Last, we outline instructional implications and directions for future research, focusing on the affordances of GIFT.

Form: Different Approaches to SRL Scaffolding

Much like domain-content scaffolding, SRL scaffolding has many forms. One category of SRL support is scaffolding by demonstration. This approach includes scaffolding that demonstrates (or models) expert behaviors in the solution process. At the domain level, several ITSs offer worked examples that show the required steps to solve a given problem (Salden, Aleven, Renkl & Schwonke, 2009). A similar approach could be applied at the SRL level, where the ITSs demonstrate what productive SRL behaviors look like. For example, the Adaptive Peer Tutoring Assistant supports students who work in dyads of tutor-tutee. This ITS assists the tutor-student by suggesting what the ITS would have done in a similar situation to support the tutee-student (Walker, Rummel & Koedinger, 2014). Indeed, these adaptive recommendations were found to support learning better than non-adaptive versions of the environment (Walker et al., 2014). Another example comes from Crystal Island, a narrative-based game for scientific inquiry. Characters in the game, as well as embedded resources (such as books and posters) model the scientific inquiry process, to assist students in conducting their own inquiry (Lester et al., 2013). A final example for supporting by demonstrating comes from Betty's Brain, an inquiry environment that uses concept maps in a variety of topics. In this environment, Betty, a virtual student agent, demonstrates good reflective behaviors while evaluating concept maps. This support was found to improve students' own reflective behaviors (Jeong & Biswas, 2008). Overall, offering worked-examples at the SRL level seems to be an effective scaffolding strategy. However, more research is required to understand how students follow these demonstrations, and what cognitive load is associated with examples at the SRL level. Other interesting questions address the format of SRL demonstrations: should SRL demonstrations be embedded in the learning environment or provided by a pedagogical agent? What should the role of the agent be? (cf. Azavedo et al., 2012).

A second (and more common) form of SRL scaffolding is SRL prompts. These are elements of the activity that instruct students to use specific SRL strategies. SRL prompts are analogous to domain-level prompts that guide students' practice by explicitly telling them what to do. For example, an ITS can instruct students to self-explain their reasoning (Conati & VanLehn, 2000; Hausmann & VanLehn, 2007). Using SRL prompts was shown to help students reflect on their knowledge and acquire better conceptual understanding of the domain in problem-solving environments (Aleven & Koedinger, 2002), scientific inquiry activities (Holmes, Dan, Park, Bonn & Roll, 2014), hypermedia environments (Azevedo et al., 2012), and games (Lester et al., 2013). Prompts ensure that students apply and practice the desired SRL strategies. In addition, many ITSs offer feedback on students' responses to these prompts. This feedback helps students learn the domain and possibly reflect on their use of SRL skills.[1] One of the disadvantages of prompts, however, is lack of learner control. Since the system chooses the time and strategy for students to apply, prompts may reduce students' agency in making those decisions. Like other forms of scaffolding, it is of interest to evaluate how SRL prompts should be faded to better support learning (as done in domain-level scaffolding, cf. Salden et al., 2009). It is possible therefore that prompts, although they tend to improve domain-level learning while they are in effect, may not improve learners' SRL when they are taken away. Ideally, learners would have learned to prompt themselves or to pause and reflect at

---

[1] SRL feedback is often not feasible. To offer SRL feedback, the environment should constraint students' responses to simple interactions (e.g., using dropdown menus), and it should be able to identify unproductive SRL behaviors. Thus, offering feedback in exploratory activities, where desired behaviors and ways of expressing them are rich and less defined, remains a challenging task.

appropriate times, without the support. However, the extent to which this happens has not been studied, to the best of our knowledge.

A third family of scaffolding techniques is cognitive tools. Cognitive tools are components in the environment that help students offload cognitive processes associated with the task (Jonassen, 1992). These tools help students carry out complex processes such as organizing their exploration (van Joolingen, 1998). For example, in Smithtown, a discovery-learning environment for economics, students are given a tool to facilitate hypothesizing and predicting (Shute & Glaser, 1990). This hypothesis builder includes domain concepts such as supply and surplus, and it helps students express potential causal relationships between these. Thus, while the tool includes domain-level concepts, the concepts lack semantic meaning, and the focus of the tool is on how to hypothesize and how hypotheses drive experimentation. Similar hypothesis tools are used across a variety of scientific domains (e.g., de Jong, 2006; Gobert, Sao Pedro, Raziuddin & Baker, 2013; Lester et al., 2013; Veermans, de Jong & van Joolingen, 2000). Cognitive tools may allow for greater student agency, as the student decides if and how they will be used. Cognitive tools may also incorporate feedback, which we discuss next. The use of these (meta)cognitive tools was shown to improve the learning process (Manlove, Lazonder & de Jong, 2007). However, their relationship with domain-level learning is not always clear, as discussed further below.

A fourth category of SRL scaffolding is feedback. Domain-level feedback focuses on the correctness of students' answers, while SRL feedback focuses on students' strategy use. To be considered SRL feedback, the feedback should be given on students' choice of actions, rather than their content. For instance, Scooter the Tutor is an animated dog that becomes angry when students seem to be gaming the system; that is, attempting to progress within the activity by taking advantage of features of the environment (Baker et al., 2013). SRL feedback can also be domain-independent. For example, the Help Tutor is an automated tutor agent that gives students explicit feedback on their choices to seek or avoid help (Roll, Aleven, McLaren & Koedinger, 2011b). Messages from the Help Tutor include no domain-specific information (e.g., "take your time and read the hint carefully" or "Consider trying to solve this step without another hint. You should be able to"). It is important to emphasize that even domain-independent feedback is trigged by domain-specific behaviors. As desired behaviors are contingent on students' prior knowledge and experiences, SRL feedback should adapt to students' proficiencies as applied to the specific task. SRL feedback oftentimes gives information about the desired application of the strategy. For example, Smithtown, described above, and Science Learning by Inquiry (SLINQ) require students to engage in effective experimentation. If students design confounded experiments to the extent that they hinder their domain learning, the systems recommend that they change only one variable at time (Gobert et al., 2013; Shute & Glaser, 1996).

One approach with promising but limited empirical support is grounded feedback. Grounded feedback creates a link between the to-be-learned content and a familiar representation that is grounded in the student's prior knowledge. For example, Darts is a game in which students need to guess the numerical value of a target on a number line (Dugdale, 1992). The game provides grounded feedback as students' numerical guesses, even when they miss the target, are plotted on the same number line, thus encouraging the students to reflect on the relative magnitudes of their different guesses. Grounded feedback is hypothesized to support triangulation: the student recognizes the correct or incorrect application of a to-be-learned skill by evaluating the outcomes of their actions using an alternative, familiar representation (which could be situational, e.g., Nathan, 1998; visual, e.g., Dugdale, 1992; or based on already-mastered procedures, e.g., Mathan & Koedinger 2005; Roll et al., 2010). Furthermore, the student can often extract additional information that can assist them in improving their answers.

One example highlighting the potential of this approach can be seen in ANIMATE (Nathan, 1998). ANIMATE is an environment for learning to translate story problems into algebraic expressions. In ANIMATE, students' equations drive an animation of the characters in the story problem. By evaluating

the match between the animation and the story, students can tell if their equations reflect the problem accurately. Students who received this grounded feedback performed better on a post-test, compared with students who were given only immediate correctness feedback on their equations. Another interesting environment that uses a similar form of grounded feedback is Alice (Cooper, Dann & Pausch, 2000), where students learn to program by creating stories and seeing these stories played out. A different form of grounded feedback can be found in the Invention Support Environment. This environment is an ITS that asks students to develop their own methods for calculating different statistical concepts (such as variability or weighted average), prior to learning the canonical solutions. Grounded feedback in this environment is given by including sets of contrasting cases that highlight deep features of the target domain (Schwartz, Sears & Chang, 2007). Applying partial methods to the contrasting cases yields results that are intuitively wrong. Students who were prompted to evaluate their methods using the contrasting cases revised their methods more often than students who received the same contrasting cases without the attention-directing prompts. Furthermore, the attention-focusing prompts also improved students' debugging ability two months after the study, even though all students practiced the taught procedures during these two months (Holmes et al., 2014). Thus, instructing students to seek grounded feedback improved their inquiry behaviors and outcomes.

The Interaction Between Domain-Level and SRL-Level Support

The relationship between domain-level and SRL-level scaffolding is interesting, and at times, they may be at odds with each other (Stampfer & Koedinger, 2013). Domain-level scaffolding may define the solution process for students (e.g., sub goals), automatize the use of learning tools (e.g., system-triggered hints), and evaluate students' performance for them (e.g., immediate feedback; Corbett & Anderson, 2001). While these characteristics are productive for learning the target topics, such scaffolding may reduce students' use of SRL strategies. Mathan & Koedinger (2005) provide one example for negotiating domain-level scaffolding (using immediate feedback) with SRL-level feedback (using grounded feedback). This work is done in a tutor for writing formulas in Excel. Problems in this tutor were designed so that errors in entering formulas in Excel lead to implausible outcomes with unreasonable magnitudes (such as $30*10 = 6,000$). Thus, students who monitored their performance could detect their own errors. To give students an opportunity to monitor their performance, system-generated feedback was postponed until after students were given a chance to detect their own errors. Students who received this self-regulated learning feedback showed greater learning and long-term transfer gains, compared with students who were given immediate domain-level feedback. Further, these students were better at troubleshooting in the post-test environment, which did not include SRL-support. The tension between domain-level and SRL-level scaffolding exists also in other forms of scaffolding, such as prompts. Schworm & Renkl (2006) evaluated an environment for instructional design using worked examples. They found that, in the presence of self-explanation prompts, offering on-demand hints hindered learning. Instead, not offering hints to learners encouraged them to self-explain their work and consequently learn better. On the other hand, other examples show that, at times, supporting greater agency on the part of students may be counter-productive for learning. Ecolab, an ITS for learning about the ecology of ponds, offers an interesting example in that regard. Different versions of the environment have different levels of scaffolding and thus different levels of learner control. An evaluation with fifth-graders found that the less control students had over their learning (that is, fewer opportunities to practice SRL strategies), the better their domain learning (Luckin & de Boulay, 1999).

While these examples show how domain-level support may hinder SRL behaviors, interestingly, SRL support may hinder domain-level learning. For example, Manlove and colleagues (2007) evaluated a discovery-learning environment with demonstrations and prompts that augmented the cognitive tools. The addition of the SRL demonstrations and prompts helped learners engage in better inquiry behaviors. For example, learners set more goals and revisited them more often. At the same time, students who received

this support learned less at the domain level. This could be due to excessive cognitive load that was introduced by the tools, due to less time to engage in actual experimentation, or due to another reason. Overall, this example demonstrates the tension that exists between domain-level and SRL-level support.

The expertise-reversal effect (Kalyuga, 2007) suggests that at the domain level, novice learners need greater support compared with experts. We speculate that similar aptitude-treatment interactions can be found also at the SRL level. When students are capable of applying the SRL strategies (as with monitoring in Mathan & Koedinger, 2005, and self-explanation in Schworm & Renkl, 2006), providing them with the agency to do so is better for learning at both levels. When students fail to apply appropriate strategies (Luckin & du Boulay, 1999), more explicit support is warranted. Initial results suggest that some SRL training may, indeed, help novice learners more than experts (Chi & VanLehn, 2010). However, more work is warranted on the effect of SRL support on learners with varied levels of expertise on the supported strategies.

Objectives: From Domain Learning to Metacognitive Learning

The examples given above show that SRL scaffolding often improves domain learning. However, can we aim higher than that? Can support for SRL achieve the ambitious goal of helping students learn to regulate their learning, and thus become more competent learners, in a manner that transfers to novel tasks, topics, and environments?

Taking a decompositional approach to SRL, we seek transfer of the same strategies that were supported. We previously proposed a hierarchy of four goals for SRL scaffolding (Koedinger et al., 2009). Within the supported environment, students should 1) apply better SRL behaviors and 2) demonstrate better domain learning. Then, in transfer activities without the SRL scaffolding, students should again demonstrate 3) better SRL behavior and 4) improvement in future domain learning. The considerable progress in the years since we first proposed this framework allows us to evaluate characteristics of SRL support that seek to improve future learning, and specifically, goal 3 (transfer of SRL skills). Studies that address this challenge are detailed in Table 1, together with a summary of their findings. We group these findings to three dimensions of transfer of SRL skills across components, topics, and environments.

Same Activity, Same Environment, New Task Components

In some cases, students transfer their SRL skills within the same activity to components of the task that do not include SRL scaffolding. For example, students who were prompted to tet oftern in a virtual lab environment continued to tets more also on later, unsupported, phases of the task (Roll, Yee & Cervantes, 2014b). It seems that transfer of the same SRL skills to unsupported components within the same activities, environments, and topics is relatively straightforward. What leads to this near transfer? It is reasonable to assume that such transfer happens not due to skill acquisition, but rather, due to adoption of certain mindsets and attitudes. Acquiring SRL skills is hard, and it is unlikely that students gain generalized, lasting SRL skills from short interventions. However, support that encourages better SRL practices in early components of an activity may trigger a mindset to use these practices that persists during subsequent components of the same activity. Applying SRL strategies is, to a large degree, a matter of work habits (Butler Cartier, Schnellert, Gagnon & Giammarino, 2011).

Table 1. Evaluating the transfer of SRL skills once SRL support is removed.

Construct	Context	Citation	SRL Scaffolding	Transfer of SRL skills		
				Transfer to new components within activity	Transfer to new activities, same environment	Transfer to new environment
Help seeking	Problem solving, geometry, high-school	Roll et al., 2011b	Feedback (domain independent help-seeking messages)	N/A	Improved	No effect
Self-assessment	Problem-solving, Algebra, middle school	Long & Aleven, 2013a	Prompts (integrated with an Open Learner Model)	N/A	N/A	No effect
Self-assessment	Problem-solving, geometry, high school	Long & Aleven, 2013b	Prompts (on paper)	N/A	Improved for low-achieving students	N/A
Self-assessment	Problem-solving, geometry, high-school	Roll et al., 2011a	Prompts (self-assessment tutor with feedback)	N/A	Improved for predicting successes	N/A
Planning / monitoring	Invention activities, statistics, college	Holmes et al., 2014	Prompts and grounded feedback (using contrasting cases in invention activities)	Improved	No effect	N/A
Planning / monitoring	Learning by teaching, ecology, elementary school	Biswas et al., 2008	Prompts and modeling (a reflective teachable agent)	N/A	Improved	N/A
Planning / monitoring	Microworlds, physics, college	Roll et al., 2014b	Prompts (reflective questions on paper)	Improved	N/A	N/A

At times, students may have the desired SRL skills, but may not see the need or benefits of applying them, contingent on their perception of the task requirements. An interesting example in that regard is the early work on self-explanation (Chi, De Leeuw, Chiu & LaVancher, 1994). Students who were prompted to self-explain learned better, even though they received no feedback on their explanations and they were not taught how to self-explain. Thus, it was merely asking students to apply an SRL skill that improved their learning. It may be that certain work habits that are facilitated by the scaffolding prime the use of certain SRL strategies. We demonstrate this adoption of SRL mindsets with the PhET D/C Circuit Construction Kit. This environment invites students to learn about D/C circuits by building electric circuits and measuring their attributes (voltage, current, etc.). We evaluated students' transfer of inquiry behaviors and attitudes when transitioning from a highly scaffolded activity to a minimally scaffolded activity within the same environment and during a single session. Students who received a combination of domain-level and SRL-level prompts transferred the prompted behaviors to the unsupported, yet related, activity (Roll et al., 2014b). Trying to explain this transfer, we surveyed students' attitudes and beliefs about the activity. Students who received a highly scaffolded activity adopted certain beliefs about the goals of the activity and the value of different strategies, and transferred these attitudes to the later activity (Roll, Yee & Cervantes, accepted-a). Thus, it is likely that this near transfer of SRL behaviors is the outcome of adopting the mindsets of good inquiry rather than learning new skills.

New Topics, Same Environment

A second kind of SRL transfer looks at students' strategies while learning new topics within the same environment, once the SRL support is removed. One example of this transfer is found in Betty's Brain, an ITS in which students create concept maps in a variety of topics (Leelawong & Biswas, 2008). In one study, students were assigned to either a Learning-by-Teaching condition (in which students learned by teaching a virtual agent) or an SRL condition (in which the virtual agent reflected on the map; Jeong & Biswas, 2008). Students who received the SRL support were more likely to trace inferences made by their maps in a transfer topic, when no reflection prompts were given. In our work on help-seeking (Roll et al., 2011a), students were given adaptive feedback on their help-seeking actions on the topics of angles and quadrilaterals in a geometry tutor. Students who received the feedback transferred better help-seeking skills to a variety of new topics within the same ITS (which included the same help resources and overall look-and-feel), even when no support was offered. Interestingly, students did not transfer across topics after receiving SRL feedback within a single topic (i.e., angles). Rather, transfer was found only after students received the SRL feedback across two different topics. It may be that initial encoding of the feedback was too topic-specific, while being exposed to the same SRL feedback across topics helped students extract its topic-independent nature and perhaps even its domain-independent nature. Additional examples show that this level of transfer is hard to achieve. For example, in our work on supporting inquiry, mentioned above, students transferred their improved behaviors to unsupported components within the same activities, but did not transfer their improved behaviors to new topics within the same learning environment (Holmes et al., 2014). In another study on the topic of self-assessment, we gave students feedback on their self-assessment attempts (Roll, Aleven & Koedinger, 2011a). Students became better at predicting when they will succeed also on new topics; however, they did not improve the accuracy of recognizing their knowledge gaps on transfer topics. Thus, while examples of SRL transfer across topics exist, we are yet to identify instructional requirements that achieve consistent results.

Same Topics, New Environment

Relatively few studies have looked into students' application of SRL skills in a transfer environment, albeit on similar topics. In fact, we are only aware of a handful of studies in which transfer was measured on a post-test, not in a new learning task. In a study on self-assessment, students were asked to estimate their knowledge level after completing each problem in a linear equation tutor with self-assessment prompts and an Open Learner Model that displays their skill mastery (Long & Aleven, 2013a). To evaluate transfer of self-assessment skills, students were also asked to assess their ability to solve the problems on the paper pre- and post-tests. The study found no improvement on self-assessment on paper from pre- to post-tests. Applying a similar approach in a study on help-seeking, we embedded hints in the paper pre- and post-tests (Roll et al., 2011a). Students in this study did not transfer their improved help-seeking skills, as demonstrated in the tutor, to the embedded hints on the paper post-test. There was one instance in which we found an effect for SRL training on self-assessment on the accuracy of students' self-assessment in the post-test, for low-achieving students (Long & Aleven, 2013b). However, in this study, the self-assessment prompts on the learning task were done on paper. Thus, while students transferred SRL skills from a learning situation to a testing situation, this was not a transfer across environments. Thus, in these examples, transfer of SRL skills was found only when the source and target environments were identical, whether used for instruction or for testing.

Overall, as seen in Table 1, there is not enough data to find a clear trend between type of support and transfer. It seems that the three dimensions of transfer that were identified above are organized according to their distance, or likelihood of transfer. Transfer across environments is the hardest to achieve, perhaps because SRL constructs that are conceptually similar require different behaviors in the different environments. For example, while asking for help in the Geometry Cognitive Tutor is done by clicking the hint button or searching the glossary, help requests in the paper assessment required students to remove

stickers that covered hints or apply freely available hints. The converse is also true, and transfer within activities seems more likely, perhaps since there are no changes to the context of the activity. This pattern emphasizes the highly situated and contextual nature of SRL learning. For example, students might see self-assessment as a necessary part of working with a tutor, but not of working on a paper, when they are not usually asked to self-assess. Butler and colleagues (2011) describe these challenges in terms of situating SRL within context. Students bring with them expectations, experiences, and attitudes that affect their learning. Thus, even when a skill is "acquired," students may not find it relevant in another environment, given a different goal, or using a different interaction style. Supporting similar SRL constructs across environments and topics may decontextualize the acquired knowledge and foster spontaneous transfer. Seeing the benefits of applying the same SRL strategies in different situations may also help establish students' beliefs of the effectiveness of these strategies. For example, within the Help Tutor, in order to overcome over-specificity in terms of domains, students transferred their behavior to a third topic only after receiving prompts in the context of two different topics within the same environment (Roll et al., 2011a).

Role: Self-, External-, and Co-Regulation

When scaffolding students' regulation within ITSs, many of the systems focus on directing students to apply prescribed strategies. In such cases, the system chooses the sub-goals and strategies for the student (e.g., using self-explanation prompts), and regulation of key elements in the learning process is done by the ITS. Azevedo refers to this approach as Externally Regulated Learning (ERL) (Azevedo, Moos, Greene, Winters & Cromley, 2008). The constructs of SRL and ERL are useful for discussing learning either from the student perspective (SRL) or the system perspective (ERL). However, these constructs are somewhat less relevant when the regulation emerges from negotiations between the student and the system (Luckin & du Boulay, 1999). A similar debate in regulation of learning in groups sparked the idea of co-regulation (Järvelä & Järvenoja, 2011, Hadwin, Järvelä & Miller, 2011). Co-regulation treats regulation as a joint, negotiated process between several stakeholders. At times, co-regulation between learners can take place in the context of using an interactive learning environment (e.g., Lajoie & Lu, 2012). However, the term co-regulation so far has focused on group work. Here, we would like to extend the use of co-regulation to capture ITSs where the learning process emerges from negotiations and interactions between the learner and the environment. Within the broader scope of interactive interfaces, the term "mixed initiative" has long been used to describe negotiations between the system and the user (Novick & Sutton, 1997).

Within ITSs, one context in which co-regulation between the system and the learner could happen is Open Learner Models (OLMs) (Bull & Kay, 2007; Long & Aleven, 2013a; Zapata-Rivera & Greer, 2002). Simple inspectable OLMs only show to the students the ITSs' estimation of their current learning status, while negotiable OLMs invite students to provide opinions about their learning progress and negotiate with the system to arrive at a shared, adjusted assessment (Bull & Kay, 2007). For instance, students could request a system-generated test to demonstrate their point of view regarding their skills, and the results of the test may influence the system's estimation on their learning status (Mabbott & Bull, 2006). A more flexible way of co-regulation may give students more direct control over their learning activities. For example, several projects explored a shared control over problem selection between the students and the system, in which the system selected the problem type for the students first (contingent on their ability), and then the students were responsible for picking a specific problem from that particular type (Corbalan, Kester & Van Merriënboer, 2008; Long & Aleven, submitted). While this form of shared control seems to improve ownership and motivation, its effects on learning are not yet clear.

Let us investigate the continuum between self-, co-, and external regulation in the context of help-seeking. The Geometry Cognitive Tutor has two help-seeking mechanisms: contextual hints, which offer several

levels of information relevant to the specific problem step, and a glossary, which is a searchable knowledge base of theorems and definitions. While these resources help learners regulate their learning, it is up to the learners to choose when and how to use these strategies (Aleven et al., 2006; Roll et al., in press). On the opposite side of this spectrum is a system that initiates help when the student struggles (cf., Luckin & du Boulay, 1999) and chooses the level of help for the student (e.g., Wood & Wood, 1999). These are instances of external regulation of learning, as the system decides for the leaner which hint is displayed and when. The learner still has some agency – for instance, the learner could ignore the given advice or could ask for more elaboration. Thus, we do not imply that instances of ERL remove all agency from the learner. However, ERL environments take a very active role in regulating the learning process, and not much negotiation takes place. The Help Tutor lies in between these two examples. Much like the system-initiated help described above, the Help Tutor also makes predictions about the right level of help to use at each moment of the learning process. However, rather than displaying the help to the learner, the Help Tutor merely advises the student on how much help to use, and the final choice is left to the learner (Roll et al., 2007; 2011b).

Applying a co-regulation lens to SRL scaffolding has several benefits. One advantage is that of agency. While the need to support students' regulation is clear (Aleven, Stahl, Schworm, Fischer & Wallace, 2003), a co-regulation approach invites students to take and maintain ownership of their learning processes. For example, Walker and colleagues (2014) offered peer-tutoring environment in which students help each other solve algebra problems. Rather than defining the interaction process for the student, the ITS offered strategies and hints without imposing them. The actual learning process is the outcome of contributions by all participating members: the ITS and the two students who engage in the learning process.

A second advantage of a co-regulation approach is that students get to practice key self-regulation skills. We previously highlighted the importance of tutoring SRL, as opposed to merely supporting it. By giving room for students to control their learning process, students get to practice key SRL skills, and possibly receive feedback on them. For example, the need for immediate feedback is clear in ITSs (Corbett & Anderson, 2001). However, immediate feedback does not give students the option to identify and diagnose their own errors. Thus, giving students an option to engage in reflective processes of error detection, together with implementation of grounded feedback, may benefit not only their domain-level knowledge, but also their use of monitoring strategies (Mathan & Koedinger, 2005). Overall, we believe that applying approaches that support co-regulation can offer a balance between supporting learning of domain knowledge and of SRL skills.

Recommendations and Future Research

The above review suggests that approaches for SRL support may be used successfully across domains, environments, and age groups. For example, SRL prompts to self-explain have been found useful in a variety of task domains (e.g., Aleven & Koedinger, 2002; Conati & VanLehn, 1998; Hausmann & VanLehn, 2009; McNamara et al., 2007). Moreover, offering parallel SRL support across activities may improve their effectiveness. In addition to supporting domain learning, reuse of the same kinds of SRL prompts across contexts may help students learn and transfer the target SRL skills. We look forward to studies that test the broad hypothesis that the same SRL support applied across multiple contexts enables students to internalize the support with beneficial effects for future learning.

This is an opportunity for a general architecture that applies similar pedagogies across domains, such as GIFT. GIFT offers an envelope for a large variety of tutoring services. Instructors can use GIFT to create tutoring environments on a variety of topics. Having a library of SRL scaffolds (such as common prompts and cognitive tools) could aid instructors in authoring environments that offer SRL support.

However, while authoring reusable support seems doable, identifying triggers for support is a much harder task. Triggers for initiating adaptive support are especially hard to author. In adaptive scaffolding, the system should detect the need for SRL support and match its level. The need for support is heavily contingent on domain knowledge and the student model. Though conceptually the two levels of scaffolding are interdependent, this research agenda requires an architecture in which metacognitive tutor agents can be implemented (and SRL support can be turned "on") without having to recreate a domain-level student model (Aleven et al., 2006). Finding a systematic way to incorporate SRL support without hindering domain-level support is an open challenge. Static support also poses its own design challenges, even though static support does not depend on students' knowledge. As described above, the interplay between support at the cognitive and SRL levels is not necessarily straightforward or predictable. Understanding when to offer SRL support in a way that will not reduce (or even augment) domain-level learning is challenging.

Another opportunity to expand GIFT is through a systematic exploration of SRL support in ITS. GIFT could become a research platform that evaluates modes of SRL support. As highlighted above, we believe that answering the following questions will take us closer to the vision of helping students acquire better learning skills:

- How reusable are different forms SRL support across topics, activities, contexts, and populations? What is the aggregate effect of reusing support? What adaptation is required to align support with context?

- What is the scope of transfer of SRL skills? What role do surface-level features (such as look-and-feel of the interface) play in transfer of SRL behaviors?

- How should SRL support adapt to student attributes? When should fading be introduced? What is the right balance between giving students agency over their learning and guiding them to apply productive behaviors?

- What are design guidelines for grounded feedback? What activities can benefit from this form of feedback?

- What is the relationship between support at the domain level and the SRL level? Which support should receive priority and when? When does support increase cognitive load and when does it decrease it?

Conclusion

Reviewing the literature on SRL tutoring within ITSs reveals the variety of forms of SRL scaffolding: demonstrations, prompts, cognitive tools, feedback, and grounded feedback. We further evaluated the effect of SRL scaffolding on transfer of SRL skills, and identified dimensions for this transfer. These results highlight the contextual nature of SRL knowledge, and achieving transfer across environments remains a challenging task. In fact, it seems that contextual similarity matters for SRL transfer more than the specific approach for scaffolding. Last, we highlight the value of supporting co-regulation of learning by building mechanisms for students to negotiate their learning and required support with the environment. We suggest that this could lead to SRL scaffolding that is more responsive to students' interactions with the environment, gives students more agency over their learning process, balance domain-level and SRL-level support, and subsequently, may lead to sustained gains to students' SRL skills and attitudes.

Finding the balance between domain-level and SRL-level scaffolding is challenging. Furthermore, the effect of both levels of scaffolding should be evaluated on both layers of learning goals (e.g., learning algebra and learning when to ask for help; learning concepts in economics and learning to raise a hypothesis). As evident in this chapter, ITSs can achieve transfer of SRL skills. As researchers and designers, we should aim to support that and be aware of the long-term effect of our environments on students' knowledge, SRL skills, and attitudes. So far, as a field, we have not done much to investigate transfer across environments. Perhaps a key will be to create support across environments. To get transfer to new environments, maybe there first needs to be support – well-aligned support – across different ITSs. The GIFT architecture seems to offer one good platform for such exploration.

Acknowledgments

References

Aleven, V. & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science*, *26*, 147–179.

Aleven, V., McLaren, B. M., Roll, I. & Koedinger, K. R. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, *16*, 101-128.

Aleven, V., Stahl, E., Schworm, S., Fischer, F. & Wallace, R. M. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, *73*(2), 277-320.

Azevedo, R. & Aleven, V. (Eds.) (2013). *International handbook of metacognition and learning technologies*. New York: Springer.

Azevedo, R., Landis, R. S., Feyzi-Behnagh, R., Duffy, M., Trevors, G., Harley, J. M. & Hossain, G. (2012). The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with metatutor. In *Intelligent tutoring systems* (pp. 212-221). Springer Berlin Heidelberg. doi:10.1007/978-3-642-30950-2_27

Azevedo, R., Moos, D., Greene, J., Winters, F. & Cromley, J. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development*, *56*(1), 45-72.

Baker, R. S. J. d., Corbett, A. T., Roll, I., Koedinger, K. R., Cocea, M., Hershkovitz, A., . . . Mathews, M. (2013). Modeling and studying gaming the system with educational data mining. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 97-115). Berlin: Springer.

Bull, S. & Kay, J. (2007). Student models that invite the learner in: The SMILI open learner modelling framework. *International Journal of Artificial Intelligence in Education, 17*(2), 89-120.

Butler, D. L., Cartier, S. C., Schnellert, L., Gagnon, F. & Giammarino, M. (2011). Secondary students' self-regulated engagement in reading: Researching self-regulation as situated in context. *Psychological Test and Assessment Modeling*, *53*, 73-105.

Chi, M. & VanLehn, K. (2010). Meta-Cognitive strategy instruction in intelligent tutoring systems: How, when, and why. *Journal of Educational Technology & Society*, *13*(1) 25-39.

Chi, M. T. H., De Leeuw, N., Chiu, M. & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*(3), 439-477.

Conati, C. & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, *11*, 398-415.

Cooper, S., Dann, W. & Pausch, R. (2000, April). Alice: a 3-D tool for introductory programming concepts. *Journal of Computing Sciences in Colleges, 15*(5), 107-116. Consortium for Computing Sciences in Colleges.

Corbalan, G., Kester, L. & Van Merriënboer, J.J.G. (2008). Selecting Learning Tasks: Effects of Adaptation and Shared Control on Efficiency and Task Involvement. *Contemporary Educational Psychology*, *33*(4), 733-756

Corbett, A. T. & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In J. Jacko, A. Sears, M. Beaudouin-Lafon & R. Jacob (Eds.), *CHI'2001 conference on human factors in computing systems* (pp. 245-252). New York: ACM Press.

de Jong, T. (2006). Scaffolds for scientific discovery learning. In J. Elen, R. E. Clark & J. Lowyck (Eds.), *Handling complexity in learning environments: Theory and research* (pp. 107-128). Emerald Group Publishing.

Dinsmore, D. L., Alexander, P. A. & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, *20*(4), 391-409.

Dugdale, S. (1992). The design of computer-based mathematics instruction. In J. H. Larkin & R. W. Chabay (Eds.), *Computer assisted instruciton and intelligent tutoring systems: Shared issues and complementary approaches* (pp. 11-45). Hillsdale, NJ: Erlbaum.

Gobert, J., Sao Pedro, M., Raziuddin, J., and Baker, R. S., (2013). From log files to assessment metrics for science inquiry using educational data mining. *Journal of the Learning Sciences, 22*(4), 521-563.

Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTRAT. *Educational Psychologist, 40*(4), 225-234.

Hadwin, A. F., Järvelä, S. & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In *Handbook of self-regulation of learning and performance* (pp. 65-84). Taylor & Francis.

Hausmann, R. G. M. & VanLehn, K. (2007). Explaining self-explaining: A contrast between content and generation. In R. Luckin, K. R. Koedinger & J. Greer (Eds.), *In proceedings of artificial intelligence in education 2007* (Vol. 158, pp. 417-424). Amsterdam, The Netherlands: IOS Press.

Holmes, N. G., Day, J., Park, A. H., Bonn, D. A., & Roll, I. (2014). Making the failure more productive: scaffolding the invention process to improve inquiry behaviours and outcomes in productive failure activities. *Instructional Science*, 42(4), 523-538 doi:10.1007/s11251-013-9300-7.

Järvelä, S. & Järvenoja, H. (2011). Socially constructed self-regulated learning and motivation regulation in collaborative learning groups. *Teachers College Record, 113*(2), 350-374.

Jeong, H. & Biswas, G. (2008). Mining student behavior models in learning-by-teaching environments. In *proceedings of Educational Data Mining* (pp. 127-136).

Jonassen, D. H. (1992). What are cognitive tools? In *Cognitive tools for learning* (pp. 1-6). Springer Berlin Heidelberg. doi:10.1007/978-3-642-77222-1_1

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509-539.

Koedinger, R., Aleven, V., Roll, I. & Baker, R. S. J. d. (2009). *In vivo* experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 383-412). New York: Routledge.

Lajoie, S. P. (2008). Metacognition, self regulation, and self-regulated learning: A rose by any other name? *Educational Psychology Review*, *20*(4), 469-475.

Lajoie, S. P. & Lu, J. (2012). Supporting collaboration with technology: Does shared cognition lead to co-regulation in medicine? *Metacognition and Learning*, *7*(1), 45-62.

Leelawong, K. & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education, 18*(3), 181-208.

Lester, J. C., Mott, B. W., Robison, J. L., Rowe, J. P. & Shores, L. R. (2013). Supporting self-regulated science learning in narrative-centered learning environments. In *International handbook of metacognition and learning technologies* (pp. 471-483). Springer New York. doi:10.1007/978-1-4419-5546-3_30

Long, Y. & Aleven, V. (2013a). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. In *Artificial intelligence in education* (pp. 219-228).

Long, Y. & Aleven, V. (2013b). Skill diaries: Improve student learning in an intelligent tutoring system with periodic self-assessment. In *Artificial intelligence in education* (pp. 249-258).

Luckin, R. & du Boulay, B. (1999). Ecolab: The development and evaluation of a vygotskian design framework. *International Journal of Artificial Intelligence in Education*, *10*(2), 198-220.

Mabbott, A. & Bull, S. (2006). Student Preferences for Editing, Persuading and Negotiating the Open Learner Model. *In M. Ikeda, K. Ashley & T-W. Chan (Eds.) Intelligent Tutoring Systems: 8th International Conference*, ITS 2006 (pp. 481-490). Berlin Heidelberg: Springer-Verlag.

Mathan, S. A. & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist*, *40*(4), 257-265. doi:doi:10.1207/s15326985ep4004_7

Mathews, M., Mitrović, T. & Thomson, D. (2008). Analysing high-level help-seeking behaviour in ITSs. In *Adaptive hypermedia and adaptive web-based systems* (pp. 312-315). Springer Berlin Heidelberg. doi:10.1007/978-3-540-70987-9_42

Manlove, S., Lazonder, A. W. & de Jong, T. (2007). Software scaffolds to promote regulation during scientific inquiry learning. *Metacognition and Learning*, *2*, 141-155.

McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I. B. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. *Reading Comprehension Strategies: Theories, Interventions, and Technologies,* 397-421.

Nathan, M. J. (1998). Knowledge and situational feedback in a learning environment for algebra story problem solving. *Interactive Learning Environments*, *5*(1), 135-159.

Novick, D. G. & Sutton, S. (1997). What is mixed-initiative interaction. In Proceedings of the AAAI spring symposium on computational models for mixed initiative interaction (pp. 114-116).

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, *31*, 459-470.

Roll, I., Aleven, V. & Koedinger, K. R. (2010). The invention lab: Using a hybrid of model tracing and constraint-based modeling to offer intelligent support in inquiry environments. In V. Aleven, J. Kay & J. Mostow (Eds.), *Proceedings of the international conference on intelligent tutoring systems* (pp. 115–124). Berlin: Springer Verlag.

Roll, I., Aleven, V. & Koedinger, K. R. (2011a). Metacognitive practice makes perfect: Improving students' self-assessment skills with an intelligent tutoring system. In G. Biswas (Ed.), Proceedings of the International conference on artificial intelligence in education (pp. 288-295). Berlin: Springer Verlag.

Roll, I., Aleven, V., McLaren, B. M. & Koedinger, K. R. (2007). Designing for metacognition - applying cognitive tutor principles to the tutoring of help seeking. *Metacognition and Learning*, *2*(2), 125-140.

Roll, I., Aleven, V., McLaren, B. M. & Koedinger, K. R. (2011b). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, *21*, 267-280.

Roll, I., Yee, N. & Cervantes, A. (accepted-a). Not a magic bullet: the effect of scaffolding on knowledge and attitudes in online simulations. Accepted for publication in the International Conference of the Learning Sciences.

Roll, I., Yee, N. & Cervantes, A. (accepted-b). Students' adaptation and transfer of strategies across levels of scaffolding in an exploratory environment. Accepted for publication in the proceedings of the International Conference on Intelligent Tutoring Systems.

Salden, R. J. C. M., Aleven, V. A., Renkl, A. & Schwonke, R. (2009). Worked examples and tutored problem solving: Redundant or synergistic forms of support? *Topics in Cognitive Science*, *1*(1), 203-213.

Schwartz, D. L., Sears, D. & Chang, J. (2007). Reconsidering prior knowledge. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 319-344). New York: Routledge.

Schworm, S. & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, *4*, 426-445.

Shute, V. J. & Glaser, R. (1990). A large-scale evaluation of an intelligent discovery world: Smithtown. *Interactive Learning Environments*, *1*, 51-77.

Stampfer, E. & Koedinger, K. R. (2013). Conceptual scaffolding to check one's procedures. In *Artificial intelligence in education* (pp. 916-919).

Tobias, S. & Duffy, T. M. (2009). *Constructivist instruction: Success or failure?* (p. 392). New York: Taylor & Francis.

Van Joolingen, W. (1998). Cognitive tools for discovery learning. *International Journal of Artificial Intelligence in Education (IJAIED)*, *10*, 385-397.

Veermans, K., de Jong, T. & van Joolingen, W. R. (2000). Promoting self-directed learning in simulation-based discovery learning environments through intelligent support. *Interactive Learning Environments*, *8*(3), 229-255.

Walker, E., Rummel, N. & Koedinger, K. R. (2014). Adaptive intelligent support to improve peer tutoring in algebra. *International Journal of Artificial Intelligence in Education*, *24*(1), 33-61. doi:10.1007/s40593-013-0001-9

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153-189). Mawah, NJ: Erlbaum.

Winne, P. H. (2006). How software technologies can improve research on learning and bolster school reform. *Educational Psychologist*, *41*(1), 5-17. doi:10.1207/s15326985ep4101_3

Wood, H. A. & Wood, D. J. (1999). Help seeking, learning, and contingent tutoring. *Computers and Education*, *33*(2), 153-169.

Zapata-Rivera, D. & Greer, J. E. (2002). Exploring various guidance mechanisms to support interaction with inspectable learner models. 6th International Conference on Intelligent Tutoring Systems, 442-452.

Zimmerman, B. J. (2010). Self-regulation involves more than metacognition: A social cognitive perspective. Educational Psychologist, 30(4), 217–221. doi:10.1207/s15326985ep3004_8

Natural Language and Discourse

X. Hu, Ed.

CHAPTER 14 – Issues Regarding the Use of Natural Language Discourse In Intelligent Tutoring Systems

Philip I. Pavlik Jr., Xiangen Hu, and Donald M. Morrison
The University of Memphis

Introduction

As the fundamental means of communication among people, discourse in natural language plays an important role in ITSs. Dialogue-based systems such as AutoTutor (e.g., Graesser, Wiemer-Hastings, Wiemer-Hastings & Kreuz, 1999) or Tactical Language and Culture Training System (TLCTS) (Johnson & Valente, 2009) provide a natural interface for the student, since discourse-based communication can be argued to be the natural modality of pedagogy, which even lay teachers (e.g., parents) use to help young people learn (Tomasello, Kruger & Ratner, 1993; Tomasello, 1999). The fundamental nature of language enabled pedagogy makes it a good starting place to consider instructional strategies since the generality of discourse may make it arguably a superset of all other methods of pedagogical interaction.

To begin this discussion, Brawner and Graesser (chapter 15) provide a review of conversational dialogue systems with the goal of introducing important basic concepts. As they explain, one of the original drivers behind this research has been the idea that computers could be surrogates for human tutors. One of the key principles they highlight is interactivity, which drives the learning. From a general perspective, we can see this interactivity results in many more interactions in dialogue-based learning compared with less interactive instruction like homework or lecture, which only allow for some discourse after long intervals. We might suppose that a greater quantity of interactions potentiates the instructional effect of discourse.

Lester, Lobene, Mott, and Rowe (chapter 16) make clear that just because you have a dialogue tutoring system doesn't mean you only have a dialogue tutoring system. Their work with aligning game features with dialogue in a virtual world (Crystal Island) presents an interesting challenge for the GIFT architecture because it reveals the need for seamless alignment of tutoring system components with the mechanics of a game scenario. This may, for example, in the case of virtual interactions in a three-dimensional world, lead to important dual coding effects, which will enhance learning beyond dialogue alone (Clark & Paivio, 1991). They urge researchers to engage in the task of determining which sorts of aligned game-like features in tutoring systems are most effective. In addition to the importance of this alignment, they focus on the need for advanced NLP research, which is similar to the following chapter, which also pushes for more complex algorithms to be put in control of dialog.

Morrison and Rus (chapter 17) provide more clarity to the lack of consistency in the terminology surrounding different sorts of instructional actions or moves. They propose a three-level taxonomy of tactics, strategies, and metastrategies and align this taxonomy with dialogue system moves. They then go on to propose that other terms for specific instructional methods or approaches should find a place in this taxonomy. Indeed, such a proposal maps well to the general proposal from the prior volume of this series, which suggested that tutoring is generally composed of an inner loop, outer loop, and curriculum loop (Pavlik Jr., Brawner, Olney & Mitrovic, 2013), which draws on an earlier two-tier model (VanLehn, 2006).

The chapter by Cai, Feng, Baer, and Graesser (chapter 18) furthers the overall conversational tutoring paradigm by introducing the trialogue (three person) conversational technique. A trialogue extends the

idea of discourse by allowing for two new modes of learning in the dialogue based tutoring system. First, there is the vicarious learning mode where the third agent is a synthetic student that serves as a model for the human learning. Vicarious learning is believed to be effective in part because it may be less threatening to learners, since feedback is observed and not received personally (Bandura, 1977; Craig, Sullins, Witherspoon & Gholson, 2006). For the more advanced student, trialogues offer the opportunity of assuming the role of the teacher for the synthetic student. This instantiates the longstanding notion that teaching itself may be instrumental for learning.

The discussion continues with work by Morrison, Nye, and Hu (chapter 19) that introduces some technical concerns in dialogue systems centered on the difficulty and procedures used in classifying and grading student responses during dialogue tutoring. These concerns are centered on the complexity of evaluating each student response relative to expectations. This chapter proposes two stages to this process, a first stage of classification, which determines the category of the response. For instance, a follow-up question by the student should be classified as a follow-up and not evaluated as an attempt to answer the question. Subsequent to this high-level classification, certain types of responses (e.g., true attempts by the student to provide an expected explanation) are further graded by using Latent Semantic Analysis or regular expressions to measure the content of relevant and irrelevant information the student has provided. Once a criterion is reached, the system moves on to new topics.

Core, Lane and Traum (chapter 20) conclude with a chapter that interweaves these issues of dialogue and strategies with a concern for how they are represented and controlled by the learner model. In particular, the issue of system-generated explanation is discussed, and how it is a non-trivial issue since the most flexible types of system-generated explanation are the most difficult to implement. They discuss an example of how these technical issues were resolved in a bilateral negotiation scenario called Stability and Support Operation (SASO). They conclude by making the point that while many systems are currently fairly simple, using mostly branching scripts, current interest and demand for more flexible dialogue systems that respond according to causal rules rather than scripts is growing, and GIFT may need to address this need going forward.

These chapters allow reflection on the GIFT project since they reveal concerns and needs of the community that will be served by the final GIFT system. In effect, they are constraints on the GIFT system that will either assure its success or cause its relegation. Besides the overarching importance they attach to strong support for dialogue tutoring generally, they provide specific characterizations of the functionality, constructs, and terminology that researchers will expect to find implemented in the GIFT system. To the extent that GIFT relabels its components in ways that do not match community terminology, it creates a barrier to use and understanding of the GIFT project.

While comprehensibility is always a barrier, once that barrier is overcome by framing GIFT in terms of the needs of dialogue tutor authors and researchers, the next bar is functionality. It seems that for a simple dialogue system, users will expect basic functionality such as speech act labeling and expectation misconception dialogues that branch to different responses. However, this functionality seems clearly a minimum going forward. Based on the concerns of both Lester et al. (chapter 16) and Core et al. (chapter 20), it seems that for advanced users and to facilitate research, GIFT developers would be wise to provide facilities to configure more flexible dialogue that may be controlled by production rules or other types of learner models. Ideally, selection algorithms for dialogue moves would be modular in the codebase, allowing researchers to more easily experiment with new NLP algorithms in the context of GIFT projects. Such facilities may help insure the longevity of the GIFT project by allowing it to track the latest developments in dialogue systems.

References

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Clark, J. & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210. doi:10.1007/BF01320076

Craig, S. D., Sullins, J., Witherspoon, A. & Gholson, B. (2006). The Deep-Level-Reasoning-Question Effect: The Role of Dialogue and Deep-Level-Reasoning Questions During Vicarious Learning. *Cognition and Instruction, 24*(4), 565-591.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P. & Kreuz, R. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research, 1*(1), 35-51.

Johnson, W. L. & Valente, A. (2009). Tactical language and culture training systems: Using AI to teach foreign languages and cultures. *AI Magazine, 30*(2), 72.

Pavlik Jr., P. I., Brawner, K. W., Olney, A. & Mitrovic, A. (2013). A Review of Learner Models Used in Intelligent Tutoring Systems In R. A. Sottilare, A. Graesser, X. Hu & H. K. Holden (Eds.), *Design Recommendations for Adaptive Intelligent Tutoring Systems Learner Modeling (Volume I)* (pp. 39-68). Army Research Labs/ University of Memphis.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

Tomasello, M., Kruger, A. C. & Ratner, H. H. (1993). Cultural learning. *Behavioral and brain sciences*, *16*(03), 495-511.

VanLehn, K. (2006). The behavior of tutoring systems.. *International Journal of Artificial Intelligence in Education, 16*(3), 227-265.

CHAPTER 15 – Natural Language, Discourse, and Conversational Dialogues within Intelligent Tutoring Systems: A Review

Keith Brawner[1] and Art Graesser[2]
[1] U.S. Army Research Laboratory; [2] University of Memphis

Abstract

Human tutors have made use of natural language during instruction for all of recorded history, with many differences in the manner of delivery (didactic, Socratic, peer interaction, etc.). While initial computer-based instructional systems were not able to make use of natural language, discourse, or dialogue-based instruction, modern ITSs have sought to integrate these various faculties. These systems draw from the large body of evidence of the success of these techniques. While the goal of this book is to produce design recommendations, this chapter has the additional goal of providing background information for other work within this section. This chapter reviews the natural divisions of dialogue-centric systems, elucidates the reasons for their creation, examines their successes, and recommends when and where one can make maximum use of these techniques.

Introduction

It is well known that one-to-one, human-to-human tutoring is extraordinarily effective, with effect sizes of expert human tutoring ranging between 0.2 and 2.0 standard deviations (called sigma) when compared with classroom learning (Bloom, 1984; VanLehn, 2011). While it makes sense that smaller class sizes are better, down to a minimum of a class size of one (Haddad, 1978), the observation of this effect begs two questions. The first question is, "Why is this form of learning interaction to be successful?" while the second is, "How can this effectiveness be scaled up?" In reference to the first question, there are a multitude of answers:

- Ability of the tutor to assess individual learning

- Ability of the tutor to tailor/customize content to the learner

- Ability to get the tutored student talking about the content

- Effects of peer learning when fellow students have similar abilities

- Grounding learning and communication processes

ITSs have implemented natural language conversational processes over the last 15 years (Graesser, VanLehn, Rosé, Jordan & Harter, 2001; Rus, D'Mello, Hu & Graesser, 2013; VanLehn et al., 2007). While these dialogue-based systems have been shown to be effective, they vary in their ability to execute some of the traditional dialogue processes and differ in their application. This chapter is not intended to be a review of all systems, their differences, and the variations in effectiveness. Reviews of AutoTutor, ITSSPOKE, Atlas, CIRCSIM, and others can be found in other papers (D'Mello & Graesser, 2013; Graesser, Keshtkar & Li, 2014). The goal of this chapter is to describe the common activities of dialogue-

centric tutors, to clarify the rationale for their invention and application, and provide recommendations to the field.

One issue that a designer of an ITS faces is what the system should be compared to. Some computer-aided learning technologies are created with the goal of replacing a textbook, encyclopedia, or Wikipedia. Other learning systems are intended to replace supplemental learning activities, such as homework drills or group interactions. ITS technologies typically have the goal of replacing or augmenting the teacher, with an individual or small class size intended for this level of optimization. Rather than didactic content simply being presented, ITS technologies attempt to emulate human tutors who perform individualized instructional actions. Human tutors represent dialogue-based actions instead of monologue-based ones.

Human tutors with varying subject matter and pedagogical expertise have been shown to improve learning (Graesser, Person & Magliano, 1995), but the incremental value of expert tutors over average tutors has yet to be established. The majority of tutors are older students, more experienced classmates, paraprofessionals, or adult volunteers, rather than highly trained instructional professionals. Cohen, Kulik, and Kulik (1982) reported that the impact of tutor training, ability, and age/grade differences on student learning were not significant, but the amount of tutoring experience was modest and difficult to project to serious tutoring over months. This research indicates that the typical human tutor would likely be an inexperienced teacher, but would nonetheless be effective in increasing student knowledge in a one-on-one situation.

To say simply that these paraprofessional and peer tutors are effective begs the question, "what do they do?" One answer is that they implement an interactive conversational dialogue approach to instruction, rather than a didactic, lecture-based, classroom teaching style. These styles of teaching can be differentiated by examining patterns of conversation and sequences of dialogue moves throughout the tutoring interaction. After a process of recording dialogue, these discourse-based instructional practices were dissected by performing analyses that segment, classify, and order speech acts within and between conversational turns (Graesser, D'Mello & Cade, 2009). Such a process can help to answer what practices predict learning gains for different categories of learners.

Best Practices of Dialogue-Based Tutoring

Cultural interactions have coevolved with the underlying biology, leading to the argument that cultural interactions may be as strong as those from biological processes (Bandura, 2011). These cultural interactions form the basis for many activities, including learning. When placing a student in a learning situation that has no biological or cultural motivation, there is limited learner interest in the activity, which has the effect of limiting the learning experience. The introduction of cultural elements into the learning process has benefits to the student, which is one of the reasons for rendering dialogue-based processes as a vehicle for introducing a cultural layer to learning material: cultural elements are introduced through the human tutor.

According to the "persona effect," life-like agent characters can seem like people even though they do not exhibit all of the emotions and personalities of humans. Animated pedagogical agents can enhance the experience of the students' learning, even when the agent itself is muted and non-expressive (Lester et al., 1997). When the agent is increasingly expressive (e.g., hand motions, facial expressions, etc.), there may be incremental pedagogical benefits (motivation, attention, etc.), which, in turn, would correlate with increased learning performance. Conversational agents can have cognitive benefits in addition to the motivational impact from depicting emotions and personality. The ability of human or computer agents to clarify, critique, explain, question, evaluate, articulate, reinforce, and justify the actions as part of interaction shows added pedagogical value over classroom-based practices (Graesser et al., 1995). Moreover, the

effectiveness of these tutorial interactions is related to their interaction-styled content, rather than superficial features of the interaction (Graesser, D'Mello & Cade, 2009). The dialogue moves of the helpful discourse includes asking of deep questions, adaptive feedback, hinting, prompting, asserting missing pieces of information, encouragement for low ability learners, and the grounding of referents in conversation to establish common ground (i.e., shared knowledge).

One of the core advantages of a dialogue-based tutor is allegedly its interactivity as a companion for discourse, following the practices noted above (Graesser, D'Mello & Cade, 2009; Graesser et al., 1995). Therefore, the obvious follow-up question is, "what do tutors do?" as a learning companion during the interaction that aids learning. The typical novice tutors within the school system are lacking in skill and deep subject matter knowledge, but are nonetheless effective. As such, it is beneficial to collect and analyze the activities of these unskilled tutors. Graesser and Person collected, transcribed, and dissected in rich detail the discourse patterns for 13 unskilled tutors, spanning over 100 hours of recorded video (Graesser & Person, 1994; Graesser et al., 1995). These analyses have indicated that tutor interactivity focuses on a few key conversational moves and discourse patterns, which can be leveraged for the construction of computer tutors.

The remainder of the chapter focuses on effects of the interactivity and dialogue-centered instruction. We identify some ways that the human tutors can improve and some advantages of a computer tutor. Attention is also given to some technical components of the dialogue-based computer tutors and how all of these parts can be integrated into a common architecture. An important goal is to propose recommendations for dialogue-based components that can be reused in any generalized architecture, and specifically, for the application of GIFT (R. A. Sottilare, Brawner, Goldberg & Holden, 2012).

Curriculum Script

One finding from the dissection of the tutoring corpora is that most human tutors have a tendency to lean toward scripted instruction rather than adaptation to the idiosyncratic problems of a student (Graesser & Person, 1994; Graesser et al., 1995). The curriculum script may be as simple as an ordered sequence of content and tasks, such as describing a formula and then giving a series of example problems in accordance with this list. The ordering may follow some principles of complexity, such as "single digit addition" prior to "multi-digit addition." The script may include a list of canned responses to typical questions. Curriculum scripts are well formed in the sense that they have specific expectations in the answer, but the expectations can be articulated, for some problems, in any order. Each expectation can be expressed in many ways, and the tutor's speech acts are dependent on what the student expresses in the dialogue history. This flexibility has been modeled in the AutoTutor conversation-based ITS (A. Graesser et al., 2012; Graesser, Wiemer-Hastings, Wiemer-Hastings & Kreuz, 1999). When the student expresses a misconception, articulates a partial answer, or encounters difficulty during the learning process, the tutor follows different paths and combinations of possibilities (potentially thousands) that depend on what the student says, rather than following a rigid sequence of speech acts. That is, the tutor pushes the agenda to get the expectations covered, but also flexibly adapts to the student following a set of if-then production rules. The interaction follows a five-step tutoring frame described in the next section.

One advantage of the curriculum script followed by a human tutor is that it can handle a broader diversity of variations than in classroom instruction. While stereotypical classroom instruction is didactic, one-directional, or populated with simple shallow questions (Dillon, 1988), the curriculum script in tutoring can handle lengthier reasoning and solutions to problems, with some dynamic modifications that are sensitive to the local student's needs or queries. These modifications allow for deeper reasoning about the content (e.g., why, how, what-if, etc.). This deeper content reasoning and space of options allows for the presentation of additional problems and examples that answer these questions and the movement to

advanced content quickly. During interactions with the curriculum script, the student has the opportunity to demonstrate and be presented with more knowledge about a subject matter, as well as deeper knowledge.

Human tutors can implement a more flexible curriculum script that is adaptive to an individual learner whereas classroom teachers have greater difficulty because of the large numbers of learners. This type of flexibility allows deviations from the idealized script, and allows learning to occur in an independent and natural fashion that is tailored to the individual student need. A generalized architecture should allow for the dynamic progression through content based on previous mastery of components in the curriculum script, with tailored content on the topics which are poorly understood, as is mentioned in the closing sections of this chapter.

Five-Step Tutoring Frame

Classroom-based interactions for most teachers have a strong tendency to follow a three-step dialogue interaction referred to as Initiation, Response, and Evaluation (Sinclair & Coulthard, 1975). This pattern begins with a teacher question, followed by a student response, followed by a teacher's evaluation of a student contribution. Tutoring dialogues, however, typically expand this to a five-step dialogue frame (Person, Graesser, Kreuz, Pomeroy & Group, 2001). These frames are illustrated below.

Classroom Dialogue Frame:

1. Teacher Question
   Why is the sky blue?

2. Student Response
   Something to do with wavelengths?

3. Teacher Evaluation
   Right.

Tutoring Dialogue Frame:

- Tutor asks the learner a question
  Why is the sky blue?

- Learner answers (frequently inaccurately)
  Something to do with wavelengths?

- Tutor gives short feedback
  Right.

- Learner and tutor work to improve answer quality (multi-turn), tutor assesses learner mastery during interactions
  Tutor: *Wavelengths have something to do with it. What elements of wavelengths matter for color?*

  Learner: *Different wavelengths have different color.*

  Tutor: *What about diffraction?*

> Learner: *Different colors have different wavelengths, which diffract differently.*
>
> Tutor: *Put it all together*
>
> Learner: *The blue light diffracts at the correct angle to be visible, while the other colors are diffracted into different directions. This also explains why sunsets are red.*
>
> Tutor: *Right!*

- Tutor: *Do you understand?*

> Student: *I think so.*
>
> Tutor: *Let's see Try this problem….*

Some tutoring systems, such as AutoTutor, have been designed to emulate a tutor in the five-step frame form of human tutoring. To implement this frame, AutoTutor was originally created with approximately a dozen dialogue moves: question, pump, prompt, prompt completions (correct answer), hint, correct hint answers, elaborations/assertions, summary, answers to student questions, slices/corrections of student misconceptions/errors, positive feedback, negative feedback, and neutral feedback (Graesser et al., 1999). Recent AutoTutor systems have been more detailed (Graesser, Conley & Olney, 2012), whereas others have narrowed down to five key dialogue acts (Wolfe et al., 2013). Regarding the latter, the five acts of questioning, hinting, prompting, correcting, and summarizing dialogue acts appear to be the minimal set of simplified components needed to provide dialogue-based instruction. These different systems have been shown to obtain learning gains of approximately 0.80 sigma (A. C. Graesser et al., 2012). Tutoring strategies with inductive support (e.g., forcing concrete articulation by the learner, short question-asking dialogues, five-step tutoring frame, etc.) have also been shown to increase learning gains so there is an open question of how the different strategies of interaction can account for the learning gains in tutoring (Heffernan & Croteau, 2004; Heffernan & Koedinger, 2000).

Expectation/Misconception Tailored Dialog, Deep Reasoning

The above specification of the curriculum script and the five-step tutoring frame captures a sizable portion of the processes that human tutors use in the process of instruction. Another part to this process is the tailoring of dialogue to the portions of content where student underperformance is noted during steps 4 and 5 of the five-step frame. The global (macroadaptive) component of this process is part of curriculum adjustments (i.e., selecting the next main question or problem to work on), whereas the local (microadaptive) level is left to specifically address problems with specific expectations or misconceptions (A. C. Graesser et al., 2012; Graesser, Hu & McNamara, 2005; Jackson & Graesser, 2007). Macroadaptive problem and content selection follows the microadaptive five-step tutoring frame until content completion.

These macro- and microadaptive processes are informed by human tutors and theories of learning that support the assumptions that encouraging students to actively construct explanations and elaborations of the learning material produces better learning than the tutor merely presenting information to students. The tutor tries to get the student to articulate good answers to difficult questions or solve difficult problems. To do so, the tutor is expecting the student to express "expectations" (i.e., correct pieces of information in a good answer) and prompts the student to do so. When the student expresses "misconceptions" (errors, bugs, flawed mental models), the tutor quickly corrects the student. This is the essence of expectation plus misconception tailored dialogue.

The above misconception/expectation dialogue supports deep reasoning and questioning about the content, and has been associated with better learning outcomes (Sullins, Craig & Graesser, 2010). Students who receive and/or ask deep reasoning questions are found to perform better on transfer learning and outcome learning tasks (Gholson et al., 2009). Expert tutors tend to ask these deep reasoning questions such as "why?", "how?", "what-if …?", and "what if not". These deep reasoning components are an important aspect of human tutoring activities.

Where are the Humans Lacking and How Can This Be Improved?

While much can be learned from the extraordinarily effective one-to-one, human-to-human tutoring, there are many ways in which it is imperfect. Human tutors are frequently novices, poorly trained, or assigned the role of being a peer tutor (Graesser & Person, 1994; Graesser et al., 1995). Although effective compared with classroom teaching, they leave room for improvement. This section identifies several potentially beneficial actions, which are rarely taken by human tutors. These actions are identified in order to make recommendations for computer tutors. When leveraged properly, they may possibly yield higher learning gains than the expert human equivalents.

Types of Instruction

The encouragement of active student learning, rapid error correction, and attention to affective characteristics are types of instruction by human tutors have shown to improve student learning (Graesser et al., 1995). Current human tutors, however, frequently overlook these strategies as part of a package of instruction. While human tutors have been shown to be effective, computer tutors may be able to be more effective when considering these added techniques. These techniques merit consideration for future dialogue-based tutoring recommendations.

One active student learning strategy occurs when it is the student who brings up a new subtopic for exploration. Such self-regulated learning rarely occurs during interactions with novice human tutors (Graesser & McNamara, 2010; Graesser & Person, 1994). These occasions primarily occur when attempting to resolve an apparent contradiction or being entirely stuck (Graesser & McMahen, 1993). Students ask approximately 27 questions per hour during tutoring, but genuine self-regulated learning questions are infrequent (Graesser & Person, 1994). The ITS encouragement of active student learning could by performed through direct manipulation. These manipulations may encourage self-regulated learning by planting contradictions, paradoxes, and arguments between agents, and have been implemented with systems that have multiple agents (D'Mello, Lehman, Pekrun & Graesser, 2014; Lehman et al., 2013).

With sophisticated pedagogical strategies the tutor uses one of a number of advanced techniques, such as Socratic Method (Rosé, Moore, VanLehn & Allbritton), reciprocal training (Palinscar & Brown, 1984), or modeling-scaffolding-fading (Van de Pol, Volman & Beishuizen, 2010). As noted above, novice human tutors have the tendency to adopt fairly rigid curriculum scripts, especially within well-structured domains, rather than more sophisticated, flexible strategies.

Another of the typical failings of human tutors is that human tutors favor rapid error correction. Immediate tutor error correction does not allow for the students to discover their own mistakes. Self-correction is a significant aspect of overall learning. The tendency of human tutors to rapidly correct errors blocks the development of important metacognitive skills (Bangert-Drowns, Kulik, Kulik & Morgan, 1991).

Many human tutors have a tendency to ignore affective and motivational aspects of learning, even though it has been encouraged by other authors (Lepper & Woolverton, 2002). The student (especially in K–12 application) is preparing for a lifetime of learning, so the cumulative effects of a motivational intervention

may be sufficient to generate future learning gains on a subject. An ideal tutor may be able to deflect negative feedback and build student confidence with their mastery of problems with increasing difficulty, but these goals are very difficult to implement and sometimes directly compete with each other. As an example, dialogue actions favoring social politeness may trump those that give direct negative feedback (Pearson, Kreuz, Zwaan & Graesser, 1995). Human tutors may be constrained in this manner whereas computer tutors are not. Affective tutoring strategies that have been shown to be effective are discussed elsewhere within this volume, in the sections on affect and instruction.

Types of Error

There are a number of situations in which a human tutor does not draw accurate conclusions about the success of the communication and learning during the interaction. There are documented the misperceptions of typical novice human tutors (Graesser, D'Mello & Person, 2009). These misperceptions include illusion of grounding, feedback accuracy, discourse alignment, student mastery, and knowledge transfer.

In the grounding problem, there is the assumption that the tutor and the student have shared knowledge about the meaning of the words and ideas expressed in the exchange. This assumption is often inaccurate because there is a large gap between what each other knows. Consider the following:

```
Tutor: "Force is a product of two items, can you name them?"
Student: "Yes, how big something is and how fast it is moving"
Tutor: "How fast something is moving was derived from its what?
    (tutor expects acceleration)
Student: "Its velocity"
    (student thinks this correct)
Tutor: "No, it's acceleration"
    (negative feedback resulting in frustration/confusion)
```

In this case, there is a lack of grounding on the tutor's intended referent for "how fast something is moving" and the referent for "what." Technically speaking, the appropriate referent for the first is velocity and for the second is acceleration. However, all four of these referring expressions may be functionally equivalent in the mind of the student and the student wants credit for saying velocity. There is a failure in the grounding of referents, which may end with student frustration. An expert human tutor can presumably diagnose a grounding problem as the collaborative construction of a solution, explanation, or answer to a question emerges. For computer tutors, this problem presents significant difficulty if the scripted nature of conversations does not have computational components that vigilantly check for grounding problems. The DeepTutor system attempts to rectify various grounding problems that are ubiquitous in the normal tutoring process (Rus et al., 2013).

The discourse alignment problem occurs when the perceived discourse function of a speech act is different for the tutor and student. This occurs when the students do not realize that they have been given help as part of a tutor's dialogue move. This problem can be difficult to reconcile in human tutoring, but can be easily solved in a computer system through color-coding or other interface design. Discourse misalignment occurs when the tutor gives a hint and the student doesn't realize it. The tutor may intend an assertion as a hint (e.g., "Acceleration is a change in velocity"), but the student thinks it is a mere supportive assertion rather than regarding it as a hint to give the student guidance. The solution to this problem for human and computer tutors is to be aware of the potential for miscommunication during the hint-giving process and minimize this by preceding the hint with a declaration of its discourse function (e.g., "Here's a hint. Acceleration is a change in velocity.")

The illusion of student mastery comes from a misdiagnosis of the true knowledge of the student and is related to the problem of misdiagnosis of knowledge transfer. As an example of this behavior, the student gives a correct set of words in a response, but does not really understand the complex idea that is needed. Novice tutors ask questions such as "do you understand?" after relaying a complex idea and take an affirmative response to indicate that the student understood all of the relayed information. An analysis of corpora suggest that expert tutors sometimes avoid making this type of mistake by asking a greater number of common ground questions (Graesser, D'Mello & Person, 2009), but the natural proclivity of conversation is not to do this troubleshooting. Computer tutors can do appropriate follow-ups to trouble-shoot possible problems in student mastery, but that's not what even expert human tutors typically do. Computers can track what students say in generative student answers as indications of true understanding. Computer tutors can also compare a student's answer to a normative sample of student answers that are graded on quality as an answer to a question. These are terrific solutions on what computers can do but it should be acknowledged that that is not what human tutors do, even expert tutors.

Regarding tutor feedback, human tutors have a tendency to give a greater amount of positive feedback than negative feedback. These actions may be either right or wrong, depending on the circumstance under which they are given. It is known that some expert tutors give significant positive comments as part of an affective style of tutoring (Lepper & Woolverton, 2002), but that is not what no-nonsense (direct feedback) accomplished tutors do (Graesser, D'Mello & Cade, 2011). It is difficult, however, for even expert human tutors to wholly avoid negative feedback, and there seems to be indication that this is a part of tutoring. Human tutors and carefully designed computer tutors can correct students on the content of what they say rather than merely giving short feedback whether they are right or wrong on a turn. In essence, tutor acts that resonate on the positive student content and assertions that try to correct student response may be better than minimal information (e.g., right/wrong) on prior contribution. The evolution of content in the exchange trumps short feedback (e.g., right/wrong). A different approach is to have two or more agents give their answers. The agents can argue, give each other feedback, and avoid blaming the human student for any deficits in their answers (D'Mello et al., 2014; Lehman et al., 2013). A student agent that mirrors what the student says can take the blame for the tutor agent's negative feedback. The human gets no blame for bad answers and credit for good answers. The purpose of this action is to boost the student's self-efficacy and preserves feedback accuracy of answers.

Where Do Computers Excel?

To state the obvious, the advantage of a computer tutoring system, even a complex one, is that it is a reliable mechanism. Computer tutors are available 24/7, can scale virtually infinitely, and can reliably follow a program of pedagogical principles. Computer tutors have infinite patience, can assign problems that are specific to student need, and have explicit control over the instruction. Control over instruction lends itself to well-designed experimentation. Furthermore, there is the well-documented success of these systems and a growing movement to leverage the dialogue-based approaches (D'Mello & Graesser, 2013; Graesser et al., 2014). However, more incisively, computer tutors have the capability of applying some of sophisticated strategies that are too difficult for humans and also overcoming some of the misperceptions and illusions of human tutors described in this section.

Technical Techniques and Component Parts

A textbook does not contain any type of individualization, whereas classroom teachers provide occasional adaptive instruction and tutoring much more. Early computer aided instruction (CAI, it was called) had conditional branching at a macro-level (Skinner, 1954). Simple branching programs (Crowder, 1959) were constructed from the linear programs to selected material based upon the answers to previous

material, in a fashion aligned with instructional best practices. This selection of an instructional frame was among the first types of adaption and among the first technical hurdles addressed by the field. Since this time, more sophisticated technical solutions have been developed at a more fine-grained level. Some of these systems involve natural language dialogue, the focus of this chapter.

One fundamental technical hurdle is a valid evaluation of the student's current level of knowledge and skill. In a dialogue-based system, students must be assessed based on their answers to the tutor's questions. These assessments use modern advances in computational linguistics to evaluate how well the students' natural language contributions match expected answers and to what extent they seem to be based on misconceptions. The feedback and dialogue moves of the tutor are triggered by these matches through production rules that are sensitive to contextual features and the dialogue history. The grain size of this adaptivity is substantially more fine-tuned and complex than CAI systems.

A useful review of dialogue-based intelligent tutoring may be found in previous publications (D'Mello & Graesser, 2013), which discuss the challenges in input transformation, speech-act classification, learner modeling, dialogue management, output rendering, and domain modeling. These functions are central to the operation of dialogue-based ITS, and must interact with each major component of a shell tutor such as GIFT. A sketch of these interactions is given in Figure 1, adapted from D'Mello and Graesser (2013).
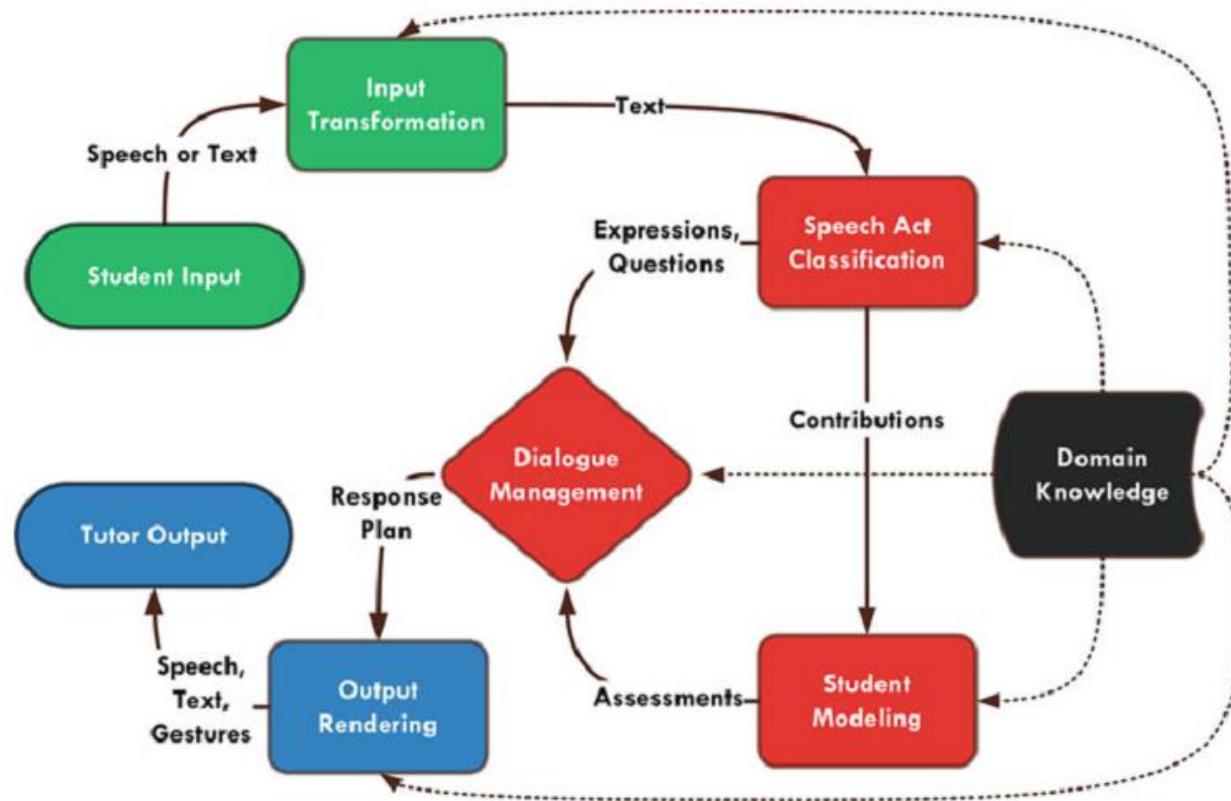
Figure 1. Interactions of various portions of dialogue-based ITS (D'Mello & Graesser, 2013).

Starting from the component of the system, which the student interacts with, there is the problem of input transformation. When input is given via keyboard, it is usually more accurate, but may have one or more additional operations performed on it. Examples of operations on text-based input include corrections for spelling (Evens et al., 1997) and the identification or modification of deeper linguistic features (Morgan, Keshtkar, Duan, Nash & Graesser, 2012). When input is spoken, there is significant challenge to process

accurately the speech-to-text translation (Seide, Li & Yu, 2011), although it is not likely that an enhancement from moderate to perfect accuracy yields any increment in learning (D'Mello, Dowell & Graesser, 2011). Consequently, text-based input is likely to be appropriate for the majority of dialogue-based tutoring tasks, assuming availability of a computer with a keyboard.

The classification of speech acts is another technical challenge. The tutor needs to respond differently to student turns that are questions, assertions, expressive evaluations, and so on. Sixteen categories of educationally relevant speech acts have been identified (Graesser & Person, 1994), but their automated detection has room for improvement. The current state of the art relies upon automatic classification based on pre-trained supervised machine learning methods such as Naïve Bayes and Decision Trees (A. Olney et al., 2003; Samei, Li, Keshtkar, Rus & Graesser, 2014). .

After speech acts have been classified, the next relevant portion of text processing evaluates the content for elements of domain content mastery. This may be done at a superficial level, such as a comparison to an ideal dialogue answer via Latent Semantic Analysis (Graesser et al., 2000; Hu, Cai, Han, Craig & Wang, 2009), an Inverse Word Frequency Weighted Overlap (D'Mello, Graesser & King, 2010), or sophisticate methods that computes logical forms (Rus, McCarthy, McNamara & Graesser, 2008). The goal of this effort is to match the student's verbal input to expectations and misconceptions and subsequently to adaptively inform further instruction. One functional question is whether an expectation has been covered, or an entire problem, well enough to progress to the next step.

Many challenges remain in dialogue-based ITSs. These include decisions when to interrupt a student, identification of when a student is on a poor line of reasoning, or what pedagogical dialogue patterns to implement in a manner that is sensitive to a learner model. In most programs, there is an overarching program of dialogue-based instruction, with sub-dialogues created, as needed, based on the subject matter competency assessments. There are open areas of research for dialogue management, with further research required in the areas of active learning and the benefits of mixed-initiative dialogues.

Areas of input transformation, speech-act classification, learner modeling, dialogue management, and domain modeling may additionally interface with secondary learning interactions. Secondary learning interactions include items such as affective states, motivation, goal orientation, and personality. Research presented elsewhere in this volume is dedicated to such subjects.

One of the emerging technical areas in dialogue-based tutoring is the ability to ask guiding questions about content. In such a scenario, the guiding question would be generated from the body of content and could be on the next item of content within the tutor's curriculum script. The ability to create an insightful question, targeted to a student's weakness, may be part of the solution to implement the sophisticated pedagogical strategies discussed earlier. Research in this area has recently begun with processes for automatic generation of concept maps (Robson, Ray & Cai, 2013), generating questions from concept maps (A. M. Olney, Graesser & Person, 2012), and ranking questions in context (Becker, Palmer, van Vuuren & Ward, 2012).

Another of the emerging technical challenges lies in the area of trialogues. While another section of this book deals with the management of affect states through pedagogy, trialogues represent a unique type of conversational interaction. The trialogue involves two characters that are able to interact with each other and with the student. These interactions can be used to instruct via their assertions and debates with each other. As an example, a tutor agent may argue with a student agent about an event ("I believe that the event has happened for these reasons"), yielding a form of instruction via example and clarification. This type of technique shows early potential for inducing confusion in the student, and additionally, shows that the student can effectively learn the various positions of the tutor (Lehman et al., 2013).

Integration Into an Architectural Paradigm

As discussed above, human tutors execute some of discourse patterns very well and it would be desirable to emulate these in an ITS. However, there are other strategies that humans do not execute, but computers are well equipped to deliver. For example, computers are better equipped to perform fine-tuned student modeling and adaptive instruction. Humans are not at all able to track such detail, perform complex mathematical computations, and generate next steps that are sensitive to the individual student's ZPD. Such detail, computation, and subtle tuning is beyond what any human could perform on the fly.

There is a third category of conversational mechanisms, which are rarely performed by expert human tutors but have the potential to yield incremental learning gains beyond the current human tutors. Tutors rarely implement sophisticated pedagogical techniques such as *bona fide* Socratic tutoring strategies (Collins, 1975), modeling-scaffolding-fading (Rogoff & Gardner, 1984), Reciprocal Teaching (Palinscar & Brown, 1984), frontier learning (Sleeman & Brown, 1982), building on prerequisites (Gagné & Gagné, 1985), or diagnosis/remediation of deep misconceptions (Lesgold, Lajoie, Bunzo & Eggan, 1988). These are briefly described below:

1. *Socratic tutoring*. The tutor asks good questions that stimulate the student to self-discover their own knowledge gaps and misconceptions, followed by a self-regulated activity of correcting their own knowledge deficits.

2. *Modeling-scaffolding-fading*. The tutor models a good strategy or skill first. Then the student actively performs it with the tutor scaffolding with correction and feedback. Then the tutor eventually fades as the student is self-sufficient.

3. *Reciprocal Teaching*. The tutor and student take turns solving a problem or answering a question, with the partner giving feedback and scaffolding good moves.

4. *Frontier learning*. The tutor presents problems that slightly extends the student's capabilities, at the edge of the ZPD.

5. *Building on prerequisites*. The tutor starts with basic building blocks of skills and builds on the prerequisite structure.

6. *Diagnosis and remediation of deep misconceptions*. The tutor identifies the deep mental models that explain the student's errors and then guides instruction to correct the misconception.

The above strategies are too complex for human tutors and for computers to implement rapidly, and instead should revolve around previously authored content. ITS technologies have attempted to achieve each of these, but with limited success or with very limited knowledge domains. One direction for future research is to make serious attempts to implement these sophisticated tutoring techniques in ITS and assess the resulting learning gains. It is conceivable that the enhanced ITS that combine these strategies with typical human tutoring strategies will reach the 2 sigma dream of Bloom (1987).

GIFT is an architecture for the support of ITSs. Systems such as this can be known as "shell tutors:" they do not tutor specific content or in a specific way, but instead enable the import of various instructional techniques and subject matter. GIFT, the eXtensible Problem-Specific Tutor (xPST) (Gilbert, Blessing & Kodavali, 2009), AutoTutor (Wiemer-Hastings et al., 1998), and the Cognitive Tutor (Anderson, Corbett, Koedinger & Pelletier, 1995) may all be considered part of a family of tutors that are architecturally agnostic to content. GIFT consists of a number of fundamental modules: the Sensor Module, the Learner Module (R. Sottilare, Graesser, Hu & Holden, 2013), the Pedagogical Module, and the Domain Module.

The Sensor and Learner Modules have the responsibility to detect various student states and traits in order to inform instructional strategy decisions. The Pedagogical Module chooses the instructional strategy (e.g., dialogue-based tutoring with scaffolding). The Domain Module contains the content and the assessments of student performance on this content. The Pedagogical Module contains a model of instruction from which to select "instructional strategies."

Through an integration with the AutoTutor framework, GIFT has begun to support dialogue-based instruction.. At the time of writing, an AutoTutor interaction supported by GIFT can assess student understanding of selected concepts. It can perform these actions as a standalone system, or as part of a video game or other learning experience. It adds the student knowledge to the Learner Module and can support interactions through the recommendations on hinting, prompting, or pumping requested by the Pedagogical Module.

The generalization of the AutoTutor approach to instruction allows AutoTutor to be used when instructionally appropriate and avoided when it is deemed best to present content directly (such as through a PowerPoint presentation) or assess content using, for example, a multiple-choice test. The ability to keep author content apart from an instructional engine allows for both the creators of content and the creators of ITSs to focus on their domain of expertise. The architectural distinction between content (Domain Module) and instruction (Pedagogical Module) allows a type of instruction to permeate through the many different training domains. As a concrete example, dialogue-based instruction is represented as an overarching pedagogical strategy, implemented with content from a specific domain. In theory, this approach allows for the rapid construction of ITSs through insulating the content author from decisions about how the content should be instructed.

Both AutoTutor and GIFT come equipped with several tools for authoring content. The combination of these tools and methods will allow a single framework to leverage the benefits from the various sets of tools, types of instruction, and types of content. An upcoming authoring advisory board and this book in this adaptive tutoring book series will help to move the field in the direction of making these systems more usable and transparent. The third advisory board, user meeting, and book on that subject (the next volume) are intended to provide guidelines on content creation for the use of the various instructional strategies mentioned above.

It is desirable that the ITS technologies of the future will exist through some combination of existing tutoring best practices and the more elaborate pedagogical mechanisms in an ITS. These practices are not observed within tutoring interactions (Graesser et al., 1995), but they are being implemented within ITS systems under development (Goldberg et al., 2012). The combination of the use of conversational dialogues within an instructional context, melded with informational instruction and practice environments, is the future direction of ITS development.

One of the projects which addresses this need is the Tools for Rapid Automated Development of Expert Models (TRADEM) (Robson et al., 2013). The TRADEM project uses content-based instruction in combination with dialogues and deep reasoning questions, built from the content automatically. These techniques are performed in concert with other instructional developments to the GIFT architecture (ARL, 2012) to enable the interweaving of pedagogy and dialogue.

In an architectural form, this automatic process of dialogue creation exists to follow the curriculum script. This curriculum script is a set of directed graphs, based upon the ordering of the content in the original documents, representing the overall script of instruction. The curriculum script is assessable (e.g., the tutor can ask intelligent questions about the items it contains) because script nodes have been linked to both content and "mini corpora" links to documents that can be found publicly on the Internet. When a content-based question is asked, the student answer can be assessed based upon the amount of matching

to the expected answer. The curriculum script, the content it presents, the questions it asks, and a smaller corpora for assessment may represent a way to create a minimal dialogue-based tutoring system that builds on other learning objects in the virtual universe. This process for automatic creation is being merged into GIFT and can logically be expanded through efforts such as the ones listed among the technical challenges.

Acknowledgments

Bibliography

Anderson, J. R., Corbett, A. T., Koedinger, K. R. & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences, 4*(2), 167-207.

ARL. (2012, 2012). Generalized Intelligent Framework for Tutoring Release Page.

Bandura, A. (2011). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology, 52*, 1 – 26.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A. & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Becker, L., Palmer, M., van Vuuren, S. & Ward, W. (2012). *Learning to tutor like a tutor: ranking questions in context.* Paper presented at the Intelligent Tutoring Systems.

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher, 13*(6), 4-16.

Cohen, P., Kulik, J. & Kulik, C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal, 19*, 237-248.

Collins, A. M. (1975). Reasoning from Incomplete Knowledge.

Crowder, N. A. (1959). Automatic tutoring by means of intrinsic programming. *Automatic teaching: The state of the art, 116*.

D'Mello, S. K., Dowell, N. & Graesser, A. (2011). Does it really matter whether students' contributions are spoken versus typed in an intelligent tutoring system with natural language? *Journal of Experimental Psychology: Applied, 17*(1), 1.

D'Mello, S. K., Graesser, A. & King, B. (2010). Toward Spoken Human–Computer Tutorial Dialogues. *Human–Computer Interaction, 25*(4), 289-323.

D'Mello, S. & Graesser, A. C. (2013). Design of Dialog-Based Intelligent Tutoring Systems to Simulate Human-to-Human Tutoring *Where Humans Meet Machines* (pp. 233-269): Springer.

D'Mello, S., Lehman, B., Pekrun, R. & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153-170.

Dillon, J. T. (1988). *Questioning and teaching: A manual of practice*: ERIC.

Evens, M. W., Chang, R. C., Lee, Y. H., Shim, L. S., Woo, C. W., Zhang, Y., . . . Rovick, A. A. (1997). *CIRCSIM-Tutor: An intelligent tutoring system using natural language dialogue.* Paper presented at the Proceedings of the fifth conference on Applied natural language processing: Descriptions of system demonstrations and videos.

Gagné, R. M. & Gagné, R. M. (1985). Conditions of learning and theory of instruction.

Gholson, B., Witherspoon, A., Morgan, B., Brittingham, J. K., Coles, R., Graesser, A. C., . . . Craig, S. D. (2009). Exploring the deep-level reasoning questions effect during vicarious learning among eighth to eleventh graders in the domains of computer literacy and Newtonian physics. *Instructional Science, 37*(5), 487-493.

Gilbert, S., Blessing, S. B. & Kodavali, S. (2009). The Extensible Problem-Specific Tutor (xPST): Evaluation of an API for Tutoring on Existing Interfaces. In V. Dimitrova, R. Mizoguchi, B. du Boulay & A. Graesser (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 707-709). Amsterdam, Netherlands: IOS Press.

Goldberg, B., Brawner, K., Sottilare, R., Tarr, R., Billings, D. R. & Malone, N. (2012). *Use of Evidence-based Strategies to Enhance the Extensibility of Adaptive Tutoring Technologies.* Paper presented at the The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Graesser, A., D'Mello, S., Hu, X., Cai, Z., Olney, A. & Morgam, B. (2012). AutoTutor. *Applied natural language processing: Identification, investigation, and resolution. Hershey, PA: IGI Global*.

Graesser, A. C., Conley, M. W. & Olney, A. (2012). Intelligent tutoring systems. *APA handbook of educational psychology. Washington, DC: American Psychological Association*.

Graesser, A. C., D'Mello, S. & Cade, W. (2009). Instruction based on tutoring. *Handbook of research on learning and instruction*, 408-426.

Graesser, A. C., D'Mello, S. & Person, N. (2009). Meta-Knowledge in Tutoring. *Handbook of metacognition in education*, 361.

Graesser, A. C., Hu, X. & McNamara, D. S. (2005). Computerized learning environments that incorporate research in discourse psychology, cognitive science, and computational linguistics. *Experimental Cognitive Psychology and its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer. Washington, DC: American Psychological Association*.

Graesser, A. C., Keshtkar, F. & Li, H. (2014). The role of natural language and discourse processing in advanced tutoring systems. *Handbook of Language and Social Psychology*.

Graesser, A. C. & McNamara, D. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist, 45*(4), 234-244.

Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104-137.

Graesser, A. C., Person, N. K. & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*(6), 495-522.

Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W. & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine, 22*(4), 39.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P. & Kreuz, R. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research, 1*(1), 35-51.

Graesser, A. C., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., Tutoring Research Group, T. R. G. & Person, N. (2000). Using latent semantic analysis to evaluate the contributions of students in AutoTutor. *Interactive Learning Environments, 8*(2), 129-147.

Haddad, W. D. (1978). Educational Effects of Class Size *World Bank Staff Working Paper, No 280*. Washington D.C: World Bank.

Heffernan, N. T. & Croteau, E. A. (2004). *Web-Based Evaluations Showing Differential Learning for Tutorial Strategies Employed by the Ms. Lindquist Tutor.* Paper presented at the Intelligent Tutoring Systems.

Heffernan, N. T. & Koedinger, K. R. (2000). *Intelligent tutoring systems are missing the tutor: Building a more strategic dialog-based tutor.* Paper presented at the Building Dialogue Systems for Tutorial Applications, Papers of the 2000 AAAI Fall Symposium.

Hu, X., Cai, Z., Han, L., Craig, S. D. & Wang, T. (2009). Autotutor Lite *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*. Amsterdam, The Netherlands: IOS Press.

Jackson, G. T. & Graesser, A. C. (2007). *Content Matters: An Investigation of Feedback Categories within an ITS.* Paper presented at the International Conference on Artificial Intelligence in Education, Los Angeles, CA.

Lehman, B., D'Mello, S., Strain, A., Mills, C., Gross, M., Dobbins, A., . . . Graesser, A. (2013). Inducing and tracking confusion with contradictions during complex learning. *International Journal of Artificial Intelligence in Education, 22*(1), 85-105.

Lepper, M. R. & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. *Improving academic achievement: Impact of psychological factors on education*, 135-158.

Lesgold, A., Lajoie, S., Bunzo, M. & Eggan, G. (1988). SHERLOCK: A coached practice environment for an electronics troubleshooting job: DTIC Document.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A. & Bhogal, R. S. (1997). The persona effect: affective impact of animated pedagogical agents *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 359–366).

Morgan, B., Keshtkar, F., Duan, Y., Nash, P. & Graesser, A. (2012). *Using state transition networks to analyze multi-party conversations in a serious game.* Paper presented at the Intelligent Tutoring Systems.

Olney, A., Louwerse, M., Matthews, E., Marineau, J., Hite-Mitchell, H. & Graesser, A. (2003). *Utterance classification in AutoTutor.* Paper presented at the Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing-Volume 2.

Olney, A. M., Graesser, A. C. & Person, N. K. (2012). Question generation from concept maps. *Dialogue & Discourse, 3*(2), 75-99.

Palinscar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and instruction, 1*(2), 117-175.

Pearson, N. K., Kreuz, R. J., Zwaan, R. A. & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and instruction, 13*(2), 161-188.

Person, N. K., Graesser, A. C., Kreuz, R. J., Pomeroy, V. & Group, T. R. (2001). Simulating human tutor dialogue moves in AutoTutor. *International Journal of Artificial Intelligence in Education, 12*, 23-39.

Robson, R., Ray, F. & Cai, Z. (2013). *Transforming Content into Dialogue-based Intelligent Tutors.* Paper presented at the The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC).

Rogoff, B. & Gardner, W. (1984). Adult guidance of cognitive development.

Rosé, C. P., Moore, J. D., VanLehn, K. & Allbritton, D. (2001). A Comparative Evaluation of Socratic versus Didactic Tutoring.

Rus, V., D'Mello, S., Hu, X. & Graesser, A. C. (2013). Recent advances in intelligent systems with conversational dialogue. *AI Magazine,* 42-54.

Samei, B., Li, H., Keshtkar, F., Rus, V. & Graesser, A. (2014). *Context-based speech act classification in intelligent tutoring systems.* Paper presented at the The Twelfth International Conference on Intelligent Tutoring Systems.

Seide, F., Li, G. & Yu, D. (2011). *Conversational Speech Transcription Using Context-Dependent Deep Neural Networks.* Paper presented at the INTERSPEECH.

Sinclair, J. M. & Coulthard, M. (1975). *Toward an analysis of discourse: The English used by teachers and pupils*: Oxford University Press London.

Skinner, B. F. (1954). *The science of learning and the art of teaching*: Cambridge, Mass, USA.

Sleeman, D. & Brown, J. S. (1982). Intelligent tutoring systems.

Sottilare, R., Graesser, A., Hu, X. & Holden, H. (2013). Design recommendations for intelligent tutoring systems.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S. & Holden, H. K. (2012). The Generalized Intelligent Framework for Tutoring (GIFT).

Sullins, J., Craig, S. D. & Graesser, A. C. (2010). The influence of modality on deep-reasoning questions. *International Journal of Learning Technology, 5*(4), 378-387.

Van de Pol, J., Volman, M. & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review, 22*(3), 271-296.

VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist, 46*(4), 197-221.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A. & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31*(1), 3-62.

Wiemer-Hastings, P., Graesser, A. C. & Harter, D. (1998). *The foundations and architecture of AutoTutor.* Paper presented at the Intelligent Tutoring Systems.

Wolfe, C. R., Widmer, C. L., Reyna, V. F., Hu, X., Cedillos, E. M., Fisher, C. R., . . . Weil, A. M. (2013). The development and analysis of tutorial dialogues in AutoTutor Lite. *Behavior research methods*, 1-14.

CHAPTER 16 – Serious Games with GIFT: Instructional Strategies, Game Design, and Natural Language in the Generalized Intelligent Framework for Tutoring

James Lester, Eleni Lobene, Bradford Mott and Jonathan Rowe
Department of Computer Science, North Carolina State University

Introduction

Recent years have seen significant progress on game-based learning. These advances include theoretical developments (Gee, 2007; Linderoth, 2012), the creation of game-based learning environments for a broad range of curricula (Johnson, 2010; Habgood & Ainsworth, 2011; Forsyth et al., 2013; Lester et al., 2014), and the emergence of immersive game-based learning technologies for both education (Hickey, Ingram-Goble & Jameson, 2009; Ketelhut, Dede, Clarke & Nelson, 2010) and training (Johnson, 2010). Recent empirical studies demonstrate that game-based learning environments can enable students to achieve learning gains in laboratory settings (Fiorella & Mayer, 2012) as well as in classroom settings (Hickey et al., 2009; Ketelhut et al., 2010; Lester et al., 2014). A pair of recent meta-analyses independently concluded that digital game technologies are often found to be more effective than traditional instructional methods in terms of cognitive outcomes, such as learning and retention (Clark, Tanner-Smith, Killingsworth & Bellamy, 2013; Wouters, Van Nimwegen, Van Oostendorp & Van der Spek, 2013). Expanding on this conclusion, Wouters et al. suggest that, "the next step is more value-added research on specific game features that determine … effectiveness" (2013, p. 262).

A key challenge for the education community is determining how to effectively integrate established instructional strategies with successful game design principles. Over the past decade, several reports have provided clear recommendations about scientifically grounded instructional strategies that can be used by teachers or implemented in computer-based learning environments (Graesser, Halpern & Hakel, 2007; Pashler et al., 2007). These instructional strategies are derived from rigorous research conducted in multiple disciplines, they have been supported by empirical studies, and they are aligned with cognitive theories about how people learn. Despite this wealth of knowledge, recommendations for how specific instructional strategies should be implemented in specific contexts, or how they should be used in combination, remain unclear.

In the case of game-based learning, there remains a dearth of theoretical and empirical work at the intersection of instructional design and game design (Isbister, Flanagan & Hash, 2010; Linehan, Kirman, Lawson & Chan, 2011). Many questions about the implementation and effectiveness of instructional strategies in game-based learning environments have not yet even been articulated, let alone answered. These questions are especially salient in the case of intelligent game-based learning environments, which are game environments that combine the adaptive pedagogical functionality of ITSs with the engaging environments of digital games. Intelligent game-based learning environments derive their effectiveness from the ability to deploy context-appropriate instructional tactics during game-based learning interactions. Game-based learning environments have grown as an educational medium over the past several years, and the need for general design principles that align instructional strategies with game design is clear. In addition, emerging architectures for ITSs, such as GIFT (Sottilare, Goldberg, Brawner & Holden, 2012), stand to benefit from an enriched understanding of how instructional strategies can be most effectively used across game-based learning environments.

Along with the need for expanded theories and empirical evidence to guide implementations of instructional strategies in game-based learning environments, fundamental advances in artificial intelligence will be necessary to realize the medium's full educational potential. In particular, NLP stands poised to serve a critical role in the implementation of instructional strategies in game-based learning environments. NLP encompasses a broad range of computational linguistics technologies, including speech synthesis and recognition, dialogue management, natural language understanding and generation, summarization, and computational models of narrative (Jurafsky & Martin, 2009). Natural language plays a central role in human instruction. For example, linguistic phenomena such as dialogue, speech understanding, and question generation are critical elements of human-to-human educational interactions (Graesser, Person & Magliano, 1995). As the research community investigates computational models of instructional strategies, NLP is also likely to play a central role. NLP holds a particularly privileged status in game-based learning environments because the rich learning interactions afforded by digital games demand sophisticated, multi-level communication capabilities only made possible by NLP. Natural language is central to many interactions with game-based learning environments, including language-based input (e.g., text, speech), human-agent dialogue, dynamically generated narratives, and believable speech by virtual agents. However, implementing robust and accurate NLP capabilities that meet the real-time performance requirements of digital games raises significant challenges to be addressed by the research community.

In this chapter, we explore the question of how theoretically and empirically grounded instructional strategies can be effectively implemented in game-based learning environments, with a focus on how NLP can play a key role in their implementation. We review recent research from the educational games literature and discuss examples of how NLP is currently being used by educational games and ITSs. To illustrate the potential synergies between game design, instructional design, and NLP, we examine several instructional strategies in CRYSTAL ISLAND, a game-based learning environment for middle school science and literacy. We outline prospective opportunities for the implementation of game-based instructional strategies in CRYSTAL ISLAND through integration with NLP functionalities. To conclude, we discuss directions for devising generalizable models of natural language-driven instructional strategies for game-based learning environments, and we identify design recommendations and research directions for game-based instructional strategy models in GIFT.

Related Research

Over the past few years, the game-based learning community has expanded efforts to conduct empirical game-based learning studies, including studies in laboratory settings (Adams, Mayer, MacNamara, Koenig & Wainess, 2012) as well as classrooms (Hickey et al., 2009; Ketelhut et al., 2010). While this has produced a wealth of evidence on the overall effectiveness of educational games, the research literature on educational game *design* remains relatively sparse. In one of the few exceptions, Isbister, Flanagan, and Hash (2010) conducted semi-structured interviews with experienced game developers to identify key design practices and themes used by professionals in their work. The interviewees described themes such as emphasizing fun as a central design value, requiring high levels of polish and well-tuned end experiences, emphasizing deep learning content rather than 'bolted on' learning materials, supporting collaboration and specialization, designing for role-playing and emotional engagement, and including affordances for exploring complex systems. In other work, Linehan and colleagues (2011) describe methods for educational game design rooted in applied behavior analysis. Still, empirical and theoretical studies on the design of specific educational game features remain few and far between.

Notably, several intelligent game-based learning environments have begun to leverage NLP to drive core aspects of learning interactions. For example, the Tactical Language and Culture Training System (TLCTS) is a suite of story-centric, serious games designed for language and culture learning (Johnson, 2010). TLCTS uses a combination of interactive lessons and narrative scenarios to train culturally

embedded spoken and non-verbal communication skills. In another example, Operation ARIES! is a dialogue-centric intelligent tutoring system about scientific reasoning that leverages game-like features to foster student engagement and learning (Forsyth et al., 2013). Operation ARIES! combines a fantasy storyline, multimedia presentations, and three-way conversational interactions with pedagogical agents to teach students about critically evaluating research claims and understanding scientific methods. Multi-agent conversational interactions are driven by tutorial dialog and language-understanding models from AutoTutor, an ITS that has also been used for multiple domains, including computer literacy and physics (Forsyth et al., 2013).

Complementary to intelligent game-based learning environments, virtual humans draw on ITSs, game engine technologies, and NLP in order to simulate naturalistic interactions with humans within software. Over the past ten years, virtual humans have been devised for a range of education and non-education applications (Swartout et al., 2013). Virtual humans typically interact with learners through combinations of verbal and non-verbal behavior, providing advice and explanations through integrated modules for speech recognition and synthesis, natural language understanding, dialogue management, and non-verbal behavior.

Many advances in educational applications of NLP have occurred outside of digital games. Automated essay grading has been the subject of considerable interest for decades, particularly given its role in assessment and standardized testing (Valenti, Neri & Cucchiarelli, 2003). Recently, computational models of tutorial dialogue have garnered increasing interest (Boyer et al., 2011; Chi, VanLehn & Litman, 2010). Computational models of dialogue have targeted a broad range of dialogue phenomena, from low-level micro-tactics (Chi, VanLehn & Litman, 2010) to high-level tutorial strategies (Boyer et al., 2011). Moreover, dialogue models have made strides by leveraging data-driven computational frameworks such as hidden Markov models (Boyer et al., 2011) and reinforcement learning (Chi, VanLehn & Litman, 2010). ITSs for writing have begun to emerge, such as Writing Pal, which combines strategy instruction, educational games, writing practice, and formative feedback components to automatically support students' writing processes (Roscoe et al., in press). Writing Pal employs several NLP modules – including a lemmatizer, syntactic parsers, lexical databases, rhetorical analyzers, and Latent Semantic Analysis – to assess students' essays and implement formative feedback functionality. In related work, ITSs have begun to leverage fine-grained linguistic indices – including measures of lexical, syntactic, and cohesion metrics – to devise models for assessing the quality of students' written self-explanations during learning (Jackson & McNamara, 2012; McNamara et al., 2012).

Although research on NLP in game-based learning environments and ITSs shows great promise, a major gap remains in the literature concerning what role NLP should play in implementing instructional strategies in game-based learning environments. As games continue to establish themselves as an important medium for education and training, resolving this question will become critical for the success of generalizable models of intelligent tutoring such as GIFT.

Discussion

In order to begin exploring the role of NLP-driven instructional strategies in game-based learning environments, we examine the implementation and effectiveness of five categories of instructional strategies in CRYSTAL ISLAND, a game-based learning environment for middle grade science. The instructional strategies are drawn from *Lifelong Learning at Work and at Home* (Graesser, Halpern & Hakel, 2007), a report that enumerates 25 evidence-based principles of human learning that correspond to actionable instructional strategies. For the purpose of discussion, we describe instructional strategies that are currently, or planned to be, implemented in CRYSTAL ISLAND. For each instructional strategy, we discuss

how NLP should drive its implementation, what form it could take in game-based learning environments, and likely computational challenges that will arise.

CRYSTAL ISLAND Game-Based Learning Environment

Over the past several years, our lab has been developing CRYSTAL ISLAND (Figure 1), a game-based learning environment for middle school microbiology and literacy (Rowe, Shores, Mott & Lester, 2011). Designed as a supplement to classroom science instruction, CRYSTAL ISLAND's curricular focus is aligned with North Carolina Essential Standards for 8th Grade Science, we well as Common Core State Standards for reading informational texts. CRYSTAL ISLAND has served as a platform for investigating a range of intelligent tutoring functionalities, including narrative-centered tutorial planning (Lee, Rowe, Mott & Lester, in press), student goal recognition (Ha et al., 2011), and affect recognition (Sabourin, Mott & Lester, 2011). The environment has also been the subject of extensive empirical investigations of student learning and engagement (Rowe et al., 2011). Studies have indicated that students achieve significant learning gains from using CRYSTAL ISLAND, and these findings have been replicated across multiple student populations (Rowe, 2013). The latest edition of CRYSTAL ISLAND was developed with the Unity game engine, which provides 3D rendering, audio, and input device capabilities, and enables deployments in schools through web browsers.

Figure 1. CRYSTAL ISLAND narrative-centered learning environment.

CRYSTAL ISLAND features a science mystery in which students attempt to discover the identity and source of an infectious disease that is plaguing a research team stationed on a remote island. Students adopt the role of a medical field agent who has been assigned to investigate the illness and save the research team from the outbreak. Students explore the research camp from a first-person viewpoint, gather information about patient symptoms and relevant diseases, form and test hypotheses about the infection, and record their findings in a diagnosis worksheet. The mystery is solved by uncovering details about the spreading infection, testing potential transmission sources of the disease in a virtual laboratory, recording a diagnosis and treatment plan, and presenting the findings to the camp nurse.

Implementing Game-Based Instructional Strategies with Natural Language Processing

To illustrate how evidence-based instructional strategies and game design principles can be aligned, we examine five cognitive principles of learning from the perspective of CRYSTAL ISLAND: 1) stories and example cases, 2) dual code and multimedia effects, 3) organization effects, 4) explanation effects, and 5) feedback effects. We discuss instructional strategies that are built upon these learning principles, and explore how NLP can serve a critical role in realizing the strategies' full pedagogical potential in CRYSTAL ISLAND, as well as game-based learning environments in general.

Stories and Example Cases. Stories provide a natural structure for encoding experiential knowledge, and they are an integral component in meaning making (Bruner, 1991). Graesser and Ottati (1996) argue that "stories have a privileged status in the cognitive system," citing experimental findings that suggest readers process narrative texts more quickly and recall narrative information more readily than expository forms. In narrative-centered learning environments – which are a class of educational games that tightly integrate gameplay, stories, and educational subject matter – students have the opportunity to serve as active participants in dynamically generated interactive narratives (Rowe et al., 2011). Narrative-centered learning environments demand use of computational models of narrative generation, which automatically reason about plots and discourse to dynamically construct coherent and engaging plots that unfold in either text-based or 3D virtual environments (Zook et al., 2012; Lee et al., in press). Recent years have witnessed growing interest in computational models of narrative for a range of education and training applications (Si, Marsella & Pynadath, 2005; Lee et al., in press). In CRYSTAL ISLAND, data-driven models of narrative-centered tutorial planning, which integrate tutorial planning and interactive narrative generation functionalities, have yielded promising results for enhancing students' learning outcomes and problem-solving processes (Lee et al., in press; Rowe, 2013). Care must be exercised in designing interactive narratives in order to avoid harmful effects of seductive details (Rowe et al., 2009), but there are also reasons to believe that interactive narratives create opportunities for supporting emotion self-regulation processes, at least for some students (Sabourin et al., 2013). Research in this area is still in its nascent stages; a majority of computational models of narrative are investigated in only a single narrative domain and educational context. Continued research on generalizable models of real-time narrative generation will be important for leveraging the instructional promise of stories and example cases in game-based learning environments, so that they can be dynamically tailored to individual learners.

Dual Code and Multimedia Effects. Dual code and multimedia effects suggest that rich representations of educational content that leverage multiple channels in a principled manner, including both verbal and visual forms, are more effective than presentations involving only a single medium (Mayer, 2009). Game-based learning environments make wide use of multi-channel interfaces, both for input and feedback. For example, TLCTS uses simultaneous text and speech in culturally situated conversational interactions with virtual agents (Johnson, 2010). Operation ARIES! leverages models of tutorial dialogue to teach scientific reasoning skills through the medium of conversational trialogues (Forsyth et al., 2013). In CRYSTAL ISLAND, science concepts are presented in three primary formats: 1) dialogue-based interactions with virtual characters that combine text and speech, 2) graphical posters that combine high-resolution images and text-based summaries of microbiology concepts, and 3) complex informational texts that appear as books and research articles in the virtual environment. These examples include both language that is procedurally generated, as well as language that is hand-authored. In order to create generalizable instructional models that adaptively tailor multimedia presentations to individual learners, devising NLP models for speech understanding and synthesis, dialogue management, text summarization, discourse understanding, and natural language generation will be essential.

Organization Effects. Organization effects suggest that outlining, integrating, and synthesizing information can enhance students' learning outcomes (Graesser, Halpern & Hakel, 2007). A number of game-

based learning environments, including CRYSTAL ISLAND, scaffold organization processes using embedded *graphic organizers*. Graphic organizers provide visual representations of how concepts are related and text is structured (Bromley, Irwin-DeVitis & Modlo, 1995). In CRYSTAL ISLAND, graphic organizers are used to scaffold students' reading comprehension processes as they read complex informational texts about microbiology concepts. Specifically, students fill out concept matrices to record key pieces of information encountered in the informational texts (Rowe, Lobene, Mott & Lester, 2013). Completing a concept matrix involves clicking on blank cells within a matrix (i.e., table) and selecting responses from drop-down menus. After filling out a concept matrix, students can press an on-screen "Submit" button to receive immediate feedback on their responses.

In the current version of CRYSTAL ISLAND, completing a concept matrix is menu-driven; students do not generate the concept matrices themselves or construct their own responses. However, increasing the role of generative processing – such as students creating their own concept matrices – is an important direction for future work. Generative learning processes have been demonstrated to be effective for enhancing reading comprehension (Linden & Wittrock, 1981), but automatically assessing student-generated concept matrices raises significant computational challenges. Providing context-sensitive feedback on student-generated content in concept matrices requires robust natural language understanding capabilities to interpret and model students' responses, as well as understand the content of associated complex informational texts. Furthermore, natural language generation would be necessary to deliver tailored feedback about students' self-generated content. In CRYSTAL ISLAND, feedback on students' concept matrices arrives in the form of virtual text messages shown on an in-game smartphone; generating brief text messages that specifically respond to the strengths and weaknesses of students' completed concept matrices would likely require automated natural language generation facilities. While intermediate solutions are possible (Rowe et al., 2013), computational challenges in providing tailored feedback on student-generated graphic organizers will shape the extent to which educational game designers can leverage generative organization effects in game-based learning environments.

Explanation Effects. Explanation effects indicate that students benefit more from generating self-explanations of mental models than memorization of shallow facts (Fonseca & Chi, 2011). While self-explanation is highly effective for learning, care should be taken in deploying self-explanation activities during game-based learning, due to the potential risks of disrupting flow during gameplay. In CRYSTAL ISLAND, self-explanation is encouraged by a diagnosis worksheet where students record their findings and conclusions as they investigate the mystery. The diagnosis worksheet is itself a graphic organizer for students' explanations of their diagnostic problem-solving processes. It includes sections for recording patients' symptoms, laboratory test results, hypotheses, and final conclusions. Prior empirical work investigating how students complete CRYSTAL ISLAND's diagnosis worksheet indicated that maintaining a thorough, accurate worksheet is significantly predictive of learning outcomes ($p < 0.001$), particularly for students with low prior domain-knowledge (Shores, 2010).

Although these findings are promising, several directions remain for enhancing the diagnosis worksheet's efficacy. Currently, the diagnosis worksheet is menu-based, but in future work we plan to implement a version where students will write their own conclusions using free-form text; students will use a *diagnosis argumentation interface* to report their conclusions. Using the interface, students will write scientific arguments to support their diagnoses, cite supporting evidence for their claims, and describe chains of deductive reasoning. With this implementation, computational models for argumentation mining – which aim to automatically detect, classify, and structure arguments in text – are likely to serve an important role in assessing the quality and correctness of students' diagnoses (Mochales & Moens, 2011).

Feedback Effects. Providing feedback on students' task performance is an important instructional strategy, and it is also a major tenet of effective game design (Schell, 2008). In game-based learning environments, feedback comprises one half of the *loop of interaction,* which refers to the continuous

cycle of information flowing between the student and the game during gameplay. Feedback has two primary roles in game-based learning. First, feedback enables students to understand the effects of their actions on the game's virtual environment. Second, feedback informs students of the correctness, or success, of their problem-solving actions during learning. Feedback can be immediate or delayed, formative or summative. Game-based learning environments such as CRYSTAL ISLAND make extensive use of feedback.

In CRYSTAL ISLAND, students receive feedback on the effects of their actions in the virtual world, the outcomes of laboratory tests they run on scientific equipment, and the correctness of their proposed diagnosis and treatment plan when they attempt to solve the mystery. However, most of this feedback is pre-specified (i.e., canned), and it is tightly coupled to the particular action the student just performed. In contrast, feedback can be adaptively tailored based on the context in which it is delivered (Serge, Priest, Durlach & Johnson, 2013). In games, feedback is often presented from virtual characters. Automatically generating context-sensitive feedback about students' problem solving demands computational models of natural language generation to drive characters' responses. Without flexible natural language generation facilities, feedback from virtual characters is likely to be limited in its context-sensitivity, as well as limited in usefulness to learners. Furthermore, it is important that natural language generation facilities use knowledge about affective and social dimensions of feedback – such as the politeness effect (Wang et al., 2008) – in order to achieve optimal learning outcomes. By acting politely and empathetically while engaging students in natural language dialogues, virtual characters are better positioned to enrich affective dimensions of learning alongside cognitive dimensions.

Recommendations and Future Research

In this chapter, we have examined the alignment of empirically based principles of instructional strategies with game design principles. While recent research has indicated game-based learning environments hold considerable promise for a broad range of education and training settings, important questions remain about how educational games should be designed and what features are most responsible for learning effectiveness. Using the CRYSTAL ISLAND game-based learning environment as an example, we have argued that evidence-based instructional strategies can be synergistically aligned with game designs, but in order to realize their full potential, advances in NLP technologies are critical. Recent work has begun to investigate NLP in intelligent game-based learning environments, and significantly expanding this line of research is an important and promising future direction.

To address these questions, we envision a three-pronged research agenda focused on the design and implementation of instructional strategies in game-based learning environments. First, we propose systematically investigating how instructional strategies can be aligned with game design principles across a range of educational subjects, learning environments, and game genres. In some cases, games and instructional design align naturally, but in others they come into conflict. For example, there are many open questions about how to most effectively incorporate narratives into game-based learning environments for different populations, educational subjects, and settings; although narrative can enhance student interest, narratives also risk introducing seductive details. Identifying the right degrees of narrative for different types of game-based learning environment is an important question for the field. As another example, it is widely recognized that self-explanation processes are effective for learning. However, it is unclear how self-explanation should be embedded in games, particularly due to the risk of disrupting players' flow. Self-explanation often requires students to write, a skill that is rarely called for in entertainment-focused games. Identifying the role of self-explanation in game-based learning environments, and how to effectively embed self-explanation processes in games, is an important question.

Second, we recommend increased efforts toward identifying algorithmic advances in NLP that enable computational models of instructional strategies in game-based learning environments to better emulate human-level implementations of these strategies. For example, fundamental advances in computational models of dialogue will enhance the opportunities available for speech-based interfaces with games, as well as opportunities for tutorial dialogue-based interactions with virtual characters. Similarly, advances in natural language understanding will enhance intelligent game-based learning environments' capacity to assess, and provide feedback on, students' writing and explanations. Without continued progress in NLP, we are unlikely to witness intelligent game-based learning environments capable of implementing instructional strategies on par with human tutors.

Third, we propose an empirically based research program on NLP-driven instructional strategies in game-based learning environments to identify the relative effectiveness of competing techniques, and pave the way for generalized intelligent tutoring models that are highly effective, transferrable, and broadly useful. This type of research program will require deploying intelligent game-based learning environments in a broad range of settings, both inside and outside the laboratory. Furthermore, this research agenda suggests a demand for tools to implement a wide range of research study designs, as well as streamline data analysis.

While it should be noted that the specific instructional strategies discussed in this chapter focus on cognitive aspects of learning – thereby omitting important affective, motivational, and metacognitive facets – we make no claim that the specific strategies or examples cited here are comprehensive. Rather, we intend for this chapter to outline one promising path forward for enhancing the effectiveness of game-based learning environments across a broad range of subjects and educational settings.

Given these recommendations, GIFT shows particular promise as a research platform for systematic investigations of NLP-driven instructional strategy models in game-based learning environments. To further align GIFT's software infrastructure with the proposed research agenda on NLP-driven instructional strategy models for games, we suggest three potential directions. First, identifying NLP-centric requirements for inter-module messaging standards and pedagogical module designs would offer a promising first step toward implementing the necessary infrastructure for supporting the proposals laid out in this chapter. Second, providing recommendations and examples for how adaptive modules for interactive narrative generation, dialogue generation, and other NLP capabilities should be integrated with the GIFT architecture would further facilitate efforts to devise effective instructional strategy models in game-based learning environments. Finally, providing streamlined instrumentation and logging facilities for monitoring the operation of NLP-driven instructional strategy models, as well as learning processes and outcomes of students interacting with these new systems, will be critical for supporting the proposed research agenda on aligning instructional strategies and game designs in the generalized intelligent framework for tutoring.

References

Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology*, *104*, 235–249.

Boyer, K. E., Phillips, R., Ingram, A., Ha, E. Y., Wallis, M., Vouk, M. & Lester, J. C. (2011). Investigating the Relationship Between Dialogue Structure and Tutoring Effectiveness: A Hidden Markov Modeling Approach. *International Journal of Artificial Intelligence in Education*, *21*(1), 65–81.

Bromley, K., Irwin-DeVitis, L. & Modlo, M. (1995). *Graphic organizers: Visual strategies for active learning.* New York: Scholastic.

Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, *18*(1), 1–21.

Chi, M., VanLehn, K. & Litman, D. (2010). Do Micro-Level Tutorial Decisions Matter: Applying Reinforcement Learning to Induce Pedagogical Tutorial Tactics. *Proceedings of the 10th International Conference on Intelligent Tutoring Systems* (pp. 224-234).

Clark, D., Tanner-Smith, E., Killingsworth, S., Bellamy, S. (2013). *Digital Games for Learning: A Systematic Review and Meta-Analysis (Executive Summary)*. Menlo Park, CA: SRI International.

Fiorella, L. & Mayer, R. E. (2012). Paper-based aids for learning with a computer-based game. *Journal of Educational Psychology*, *104*(4), 1074–1082.

Fonseca, B. & Chi, M.T.H. (2011). The self-explanation effect: A constructive learning activity. In Mayer, R. & Alexander, P. (Eds.), *The Handbook of Research on Learning and Instruction*. New York: Routledge Press.

Forsyth, C., Graesser, A., Pavlik, P., Cai, Z., Butler, H., Halpern, D. & Millis, K. (2013). Operation ARIES!: Methods, Mystery, and Mixed Models: Discourse Features Predict Affect in a Serious Game. *Journal of Educational Data Mining*, *5*(1), 147–189.

Gee, J. P. (2007). *What video games have to teach us about learning and literacy* (2nd ed.). New York: Palgrave Macmillan.

Graesser, A.C., Halpern, D.F., Hakel, M. (2008). 25 Principles of Learning. Washington, DC: Taskforce on Lifelong Learning at Work and at Home.

Graesser, A. C., Person, N. K. & Magliano, J. P. (1995). Collaborative Dialogue Patterns in Naturalistic One-to-One Tutoring. *Applied Cognitive Psychology*, *9*(6), 495–522.

Graesser, A. & Ottati, V. (1996). Why stories? Some evidence, questions, and challenges. In R. S. Wyer (Ed.), *Knowledge and memory: The real story* (pp. 121–132). Hillsdale, NJ: Erlbaum.

Ha, E. Y., Rowe, J. P., Mott, B. W. & Lester, J. C. (2011). Goal Recognition with Markov Logic Networks for Player-Adaptive Games. *Proceedings of the 26th National Conference on Artificial Intelligence* (pp. 2113–2119).

Habgood, M. P. J. & Ainsworth, S. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences*, *20*(2), 169–206.

Hickey, D. T., Ingram-Goble, A. A. & Jameson, E. M. (2009). Designing Assessments and Assessing Designs in Virtual Educational Environments. *Journal of Science Education and Technology*, *18*(2), 187–208.

Isbister, K., Flanagan, M. and Hash, C. (2010). Designing games for learning: insights from conversations with designers. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems* (pp. 2041–2044). New York, NY.

Jackson, G.T. & McNamara, D.S. (2012). Applying NLP metrics to students' self explanations. In P.M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 261-275). Hershey, PA: IGI Global.

Johnson, W. L. (2010). Serious use of a serious game for language learning. *International Journal of Artificial Intelligence in Education*, *20*(2), 175–195.

Jurafsky, D. & Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.

Ketelhut, D., Dede, C., Clarke, J. & Nelson, B. (2010). A Multi-user virtual environment for building and assessing higher order inquiry skills in science. *British Journal of Educational Technology*.

Lee, S., Rowe, J., Mott, B. & Lester, J. (in press). A Supervised Learning Framework for Modeling Director Agent Strategies in Educational Interactive Narrative. *Transactions on Computational Intelligence and AI in Games - Special Issue on Computational Narrative and Games*.

Lester, J. C., Spires, H. A., Nietfeld, J. L., Minogue, J., Mott, B. W. & Lobene, E. V. (2014). Designing Game-based Learning Environments for Elementary Science Education: A Narrative-Centered Learning Perspective. *Information Sciences, 264*, 4–18.

Linden, M. & Wittrock, M. C. (1981). The teaching of reading comprehension according to the model of generative learning. *Reading Research Quarterly*, *17*(1), 44–57.

Linderoth, J. (2012). Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming and Virtual Worlds*, *4*(1), 45–61.

Linehan, C., Kirman, B., Lawson, S. and Chan, G. (2011). Practical, appropriate, empirically-validated guidelines for designing educational games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1979–1988).

Mayer, R. E. (2009). *Multimedia Learning*. New York: Cambridge University Press.

McNamara, D. S., Jackson, G. T., Boonthum, C., Deng, Y. & Xiangyou, H. (2012). iSTART-ME: A natural language game-enhanced comprehension strategy tutor. *Journal of South China Normal University*, 6.

Mochales, R. & Moens, M. (2011). Argumentation mining. *Artificial Intelligence and Law, 19*(1), 1–22.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M. & Metcalfe, J. (2007). *Organizing Instruction and Study to Improve Student Learning*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved February 7, 2014, from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/20072004.pdf.

Roscoe, R. D., Varner, L. K., Weston, J. L., Crossley, S. A. & McNamara, D. S. (in press). The Writing Pal intelligent tutoring system: Usability testing and development. *Computers and Composition*.

Rowe, J. (2013). Narrative-centered tutorial planning with concurrent Markov decision processes (Doctoral dissertation). North Carolina State University, Raleigh, NC.

Rowe, J. P., Lobene, E. V., Mott, B. W. & Lester, J. C. (2013). Embedded Scaffolding for Reading Comprehension in Open-Ended Narrative-Centered Learning Environments. *Proceedings of the AIED'13 Workshop on Scaffolding in Open-Ended Learning Environments (OELEs)*.

Rowe, J., McQuiggan, S., Robison, J. & Lester, J. (2009). Off-task behavior in narrative-centered learning environments. *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 99–106).

Rowe, J. P., Shores, L. R., Mott, B. W. & Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education*, *21*(2), 115–133.

Sabourin, J., Mott, B. W. & Lester, J. C. (2011). Generalizing Models of Student Affect in Game-Based Learning Environments. *Proceedings of the Workshop on Machine Learning and Affective Computing, in conjuction with the 4th International Conference on Affective Computing and Intelligent Interaction*.

Sabourin, J., Rowe, J., Mott, B. & Lester, J. (2013). Considering alternate futures to classify off-task behavior as emotion self-regulation: A supervised learning approach. *Journal of Educational Data Mining*, *5*(1), 9–38.

Schell, J. (2008). *The Art of Game Design: A Book of Lenses*. Burlington, MA: Morgan Kaufmann.

Serge, S., Priest, H., Durlach, P. and Johnson, C. (2013). The effects of static and adaptive performance feedback in game-based learning. *Computers in Human Behavior*. *29*(3), 1150–1158.

Shores, L. (2010). The Role of Cognitive and Metacognitive Tool Use in Narrative-Centered Learning Environments (Masters thesis). North Carolina State University, Raleigh, NC.

Si, M., Marsella, S. C. & Pynadath, D. V. (2005). Thespian: An Architecture for Interactive Pedagogical Drama. *Proceedings of the Twelfth International Conference on Artificial Intelligence in Education* (pp. 595–602).

Sottilare, R. A., Goldberg, B. S., Brawner, K. W. & Holden, H. K. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference.*

Swartout, W., Artstein, R., Forbell, E., Foutz, S., Lane, H. C., Lange, B., Morie, J. F., Rizzo, A. S. & Traum, D. (2013). Virtual Humans for Learning. *AI Magazine*, *34*(4), 13–30.

Valenti, S., Neri, F. & Cucchiarelli, A. (2003). An Overview of Current Research on Automated Essay Grading. *Journal of Information Technology Education*, *2*, 319–330.

Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E. & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies*, 66(2), 98-112.

Wouters, P., van Nimwegen, C., van Oostendorp, H. & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265.

Zook, A., Lee-Urban, S., Riedl, M. O., Holden, H. K., Sottilare, R. A. & Brawner., K. W. (2012). Automated Scenario Generation: Toward Tailored and Optimized Military Training in Virtual Environments. *Proceedings of the 7th International Conference on the Foundations of Digital Games* (pp. 164–171).

CHAPTER 17 – Moves, Tactics, Strategies, and Metastrategies: Defining the Nature of Human Pedagogical Interaction

Donald M. Morrison & Vasile Rus
Institute for Intelligent Systems, University of Memphis

Abstract

Along with more general terms such as *method*, *technique*, and *approach*, the terms *strategy* and *tactic* are often used interchangeably in the educational research literature, leading to potential confusion and missed opportunities for collaborative theory-building. Here we argue for more precise definitions of these terms, consistent with conventional business and military usage, in which a *strategy* is a systematic plan aimed at accomplishing a particular long-term goal, and a *tactic* is a particular action, selected opportunistically and intelligently from a set of possible actions associated with the strategy, for the purpose of attaining a near-term objective. A *metastrategy* is a higher-level plan (algorithm) for choosing from among a suite of strategies, based on changing goals and circumstances. These definitions, we argue, are useful in exploring fundamental issues regarding the nature of human teaching and learning because they force us to build instructional theory from the ground up, explaining how particular actions ("moves") produce particular results, and how the choice of one action over another is informed in some principled way. Having established this basic conceptual framework, we suggest how we are applying it to the analysis of tutorial dialogue, the primary and ancient means by which humans acquire new knowledge and skill from each other.

Introduction

In the research literature on intelligent tutoring, the terms *method*, *approach*, *technique*, *tactic*, and *strategy* get used more or less interchangeably, sometimes even within the same article. A single example, from a paper on politeness in pedagogical agents (Wang, Johnson, Rizzo, Shaw & Mayer, 2005) makes the point:

> The face threat of the instruction can be mitigated using negative politeness *tactics* [italics added], i.e., phrasing that gives the hearer the option of not following the advice, e.g., "Do you want to save the factory now?" Positive politeness *strategies* [italics added] can also be employed that emphasize common ground and cooperation between the tutor and learner, e.g., "How about if we save our factory now?"

In a recent paper titled "Instructional complexity and the science to constrain it" (Koedinger, Booth, and Klahr, 2013), the authors cite the same problem – a lamentable lack of precision in terms used to describe the various aspects of instruction in the educational research literature. "Many debates about instructional methods," they argue, "suffer from tendency to apply compelling labels to vaguely described procedures, rather than operational definitions of instructional practices" (p. 935).

However, in the same paper, readers will find that the terms *technique*, *approach*, *method*, *instructional choice*, and even *learning principle* are employed more or less interchangeably, as if these terms all refer to a single construct. For example, the authors provide a list of 30 "instructional techniques," which are listed in an accompanying table as "instructional design principles." These techniques (or principles), culled from the recent literature (Pashler et al., 2007; Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013), include an undifferentiated mix of practices and contextual factors including *spacing*, *scaffolding*,

exam expectations, *testing*, *segmenting*, and *feedback*. These constructs clearly straddle or ignore essential distinctions that would be necessary to build out the kind of science-based instructional theory that Koedinger, Booth, and Klahr call for. For example, lower-level choices such as whether to provide immediate or delayed feedback, cue prior knowledge, or prompt for self-explanation seem to represent tactical, "inner loop" choices (Vanlehn, 2006), whereas the choice to provide (or not provide) time for reflection and questioning during a lesson seems more strategic – especially given that a particular action, such as prompting for an explanation, assumes the higher-level choice of whether to provide time for reflection and questioning at all.

With these concerns in mind, we argue that discussions about the nature of human teaching and learning ought to be founded on sufficiently granular descriptions of real-time teacher-learner interactions, in which moves and counter-moves are understood as tactics, reflecting the participants' hidden strategies and metastrategies, aimed at achieving particular short-term objectives and longer-term goals, which may be aligned, or not. To be clear, we are not saying that terms such as *technique*, *method*, *approach,* and *instructional choice* ought to be avoided; rather, we are saying that it is fair and important to ask whether such terms refer to lower-level tactics or higher-level strategies, and how they might fit into a description that fully honors the transactional nature of human teaching and learning. To this end, we offer a conceptual framework for use in describing agent-based pedagogical interaction, in which the terms *move*, *tactic*, *strategy*, and *metastrategy* have particular and distinct meanings, especially as defined in relationship to each other. We then go on to explain, briefly, how we are currently using this framework for the analysis of human-human tutorial dialogue.

Toward a Transactional Understanding of Human Learning: Moves, Tactics, Strategies, and Metastrategies

We begin by defining the scope of our framework, which we define as including all instances of human learning that involve some form of real-time, intentional interaction between a learner and a "more knowledgeable other" (MKO, Vygotsky, 1930/1978). In addition to intelligent tutoring, the scope therefore includes formal classroom instruction, informal learning among peers, work-based apprentice-expert learning, and formal human-human tutoring. In order to avoid the somewhat awkward term MKO, we here use the term teacher to refer to the more knowledgeable other, asking the reader to keep in mind that the "teacher" might be an intelligent pedagogical agent, human teacher, tutor, expert, or more knowledgeable peer.

In all of these cases, at the most granular level, instruction is transactional and intentional, involving a sequence of back and forth moves by the participants. Indeed, we take the concept of a move, a particular action, as the basic building block of the framework proposed here. A move can be a physical gesture (a nod or shake of the head, pointing to an object, a physical demonstration, such as pulling a lever), a single utterance or other vocalization, or even the lack of an action where one might be expected, e.g., a "pregnant pause." In the context of a traditional ITS, a move is any single action taken by the learner, or any single system response to the learner's move. In a dialogue-based system such as AutoTutor (e.g., Graesser, Chipman, Haynes & Olney, 2005), moves are individual learner contributions and the tutor's responses, including prompts, hints, and other forms of feedback. In a serious game, moves are individual player actions and system responses.

Now, here is our first point. A *move* is only theoretically interesting if it is intentional, i.e., it is undertaken by an intelligent agent with a particular purpose in mind. A cough is just a cough, *unless* it is intended as an alert, in which case it becomes a *tactic*, i.e., an intentional move. Further, except in cases where there is only a single possible move, a tactical move represents a *choice* – in pedagogical terms, an instructional choice. Importantly, in contrast to the move itself, which is a visible action, the tactical

nature of a move is almost always hidden. If we want to understand why an ITS has given one type of response instead of another, we need access to the computer program behind it, so that we may study the algorithms that made the choice. If we are playing chess and we know something about the rules of the game, we can see that our opponent's move forces us to move our king a space to the left. We recognize the opponent's move as a tactic, but can only suspect why our opponent chose this move over some other possible move. In other words, we assume that our opponent has a plan and that the move in some way reflects the plan, but (and this is what makes a game like chess possible), the plan is hidden from us. We will call this plan a *strategy*, an algorithm or "policy" an agent uses to choose from among possible moves.

To summarize the argument to this point, we are saying that an intentional *move* is the fundamental building block for a transactional theory of human teaching and learning. In itself, a move is simply an action, which may be connected to some larger purpose, or not. Of course, most actions undertaken by humans are other intelligent agents *are* purposeful, and so, by our definition, constitute *tactics* of one form or another. However, it is possible to recognize (or at least strongly suspect) that a move is a tactic without understanding its larger purpose. It is for this reason – because tactics are visible, purposeful moves, designed to implement hidden strategies – that it is important to distinguish strictly between moves, tactics, and strategies as different constructs.

Another important thing about moves, and therefore tactics, is that moves are sequential and interactive in a way that strategies are not. For example, a strategy in the game Rock, Scissors, Paper (RSP) is to throw whatever would have lost to the opponent's last throw. If your opponent last threw scissors, you throw paper, on the belief that your opponent is more likely to make the throw that would have beaten her last one. The same strategy can be employed over the course of several turns; the determination of what to throw is therefore based on a combination of the player's selected strategy and the opponent's previous throw. A series of such throws represents a history of different visible moves (tactical choices) made in keeping with the same hidden strategy.

Now, on the basis of this history, if the other player suspects that his opponent is using the "beat last throw" strategy, he can attempt to foil it by adopting a "*repeat* last throw" strategy. This illustrates two important features of turn-taking games of this type. First, we see that choices of tactics and strategies are made in a continuously unfolding present, based partly on past moves, and partly on anticipated future moves. Second, here we see the need for a higher-level construct, which we can term a *metastrategy*. In other words, just as a tactic is a principled choice of a particular move, based on a particular strategy, so a particular metastrategy may be viewed as a principled way of choosing a strategy from among a repertoire of strategies, depending on changing circumstances. In the simple game of RSP, there are only three possible moves; however, there are multiple strategies, and assuming a move is conditioned by a strategy, the same move can represent a different tactic, conditioned by a different strategy; further, based on their assessments of the success of a given strategy as the game unfolds in real time, players can switch strategies, in accordance with a metastrategy. Over time, those with the most effective metastrategies win.

While you might argue that it is cleaner to say simply that strategies are hierarchical in nature, and that higher-level strategies can subsume lower-level ones, we find the term *metastrategy* useful because, as explained below, it allows us to talk about how tutors select from among distinct strategies such as lecturing, modeling, scaffolding, and fading (Cade, Copeland, Person & D'Mello, 2008). Theoretically, once the field has identified different metastrategies, one can imagine a system that chooses from among metastrategies, and therefore has a "meta-metastrategy," and so on. However, in practical terms, we find it both necessary and sufficient to distinguish between *strategies* and *metastrategies*.

The following table summarizes the terms we have defined to this point.

Table 1. Basic building blocks for a transactional theory of human learning.

Term	Definition	Example(s)
move	Any action that potentially alters the course of an interaction between a teacher and a learner	spoken utterance, marked pause, physical gesture, system feedback
tactic	A purposeful move, aimed at achieving a particular short-term goal, in accordance with a specific strategy [see below]	[see above]
strategy	An algorithm for selecting from a set of tactics (purposeful moves) based on current circumstances, and aimed at achieving a long-term goal	lecturing, modeling, scaffolding, fading
metastrategy	An algorithm for selecting from among a repertoire of strategies	an individual tutor's "policy" for selecting from among available strategies

Moves, Tactics, Strategies and Metastrategies in Tutorial Dialogue

Having established these definitions in general terms, as applied to any purposeful interaction between a learner and teacher, we now make a few points about the analysis of *tutorial dialogue*, a term which we mean to refer not just to one-on-one interactions between a traditional tutor and a learner, but any natural language conversation with a pedagogical intent on the part of at least one of the participants. For example, any academically oriented classroom discussion, whether formal or informal, would fall into this category, as would a work-embedded conversation between an expert and apprentice, a conversation between a mother and child, or any other such conversation with at least a partial pedagogical intent. A conversation between a learner and an intelligent tutor would also fall into this category.

Solutions to the problem of understanding the nature of these dialogues at a granular level involve developing a taxonomy of dialogue moves, then interpreting these moves as tactics connected to larger strategies. Because the strategies – the algorithms that participants use to select from among possible moves – are hidden inside the participants' heads, these strategies must be inferred by examining sequences of dialogue moves over time and making note of the conditions under which they get used, how they tend to cluster, and, ultimately, the apparent impact of different moves on the course of the dialogue, and the differences in metastrategies employed by more effective tutors compared to those employed by less effective ones.

Dialogue moves – which, in a reference to speech act theory (Austin, 1965; Searle, 1969) may also be called *dialogue acts*—include various kinds of questions, assertions, prompts, hints, expressives ("Great!), requests for confirmation ("Right?"), and confirmations ("Right..."). Given that a full discussion of the requirements of a tutorial dialogue act taxonomy would take us well beyond our page limitation, we note simply that various schemes have been developed (e.g., Graesser, Person & Magliano, 1995; Olney, Person & Graesser, 2012; Rus, Graesser, Moldovan & Niraula, 2012).

In human dialogue, as in chess, RSP, and other such games, moves are made sequentially, with each move both conditioned by the preceding move and conditioning the next. A greeting is followed by a greeting, a question by an answer, an apology by an acceptance of the apology, and so forth. Further, clusters of moves, including pairs, may be associated with higher-level constructs, which, following Cade, Copeland, Person & D'Mello (2008), we will call *modes*. Modes can represent social conventions (e.g.,

greetings, partings, praise, and so forth), or they can be pedagogical in purpose, as when a teacher asks a series of questions aimed at determining a learner's current level of skill or prior knowledge of a topic. These modes, we suggest, are associated with instructional strategies, but are not in themselves strategies. Again we can turn to the chess analogy, where different strategies are associated with different phases of the game: opening, middle, and end. Strategies that make sense in the opening phase ("Move knights and bishops from the back row…" or "Establish a solid guard for your king…") make less sense in the end game.

Finally, an important thing to keep in mind in undertaking this kind of analysis is that, assuming the participants' moves are tactical, i.e., purposeful, each participant in the dialogue is by definition acting in accordance with his or her own current strategy, which may or may not be aligned with the other's. For example, a learner's strategy might be "Pretend to understand..." while the teacher's strategy might be "Ascertain the learner's current understanding."

Figure 1 summarizes this way of thinking. Briefly, a tutorial dialogue, like any human conversation, is understood by participants and observers (eavesdroppers) as a jointly constructed, rule-governed activity. As such, participants take turns making *utterances*, through which they seek to *contribute* (Clark & Schaefer, 1989) to the conversation. These utterances may be understood as constituting a series of moves, which may be interpreted as *dialogue acts*. Since one move invites another, it often makes sense to analyze dialogue acts in pairs, called *adjacency pairs* (Sacks, 1970). For example, a question is typically followed by an answer, but can also be followed by another question.
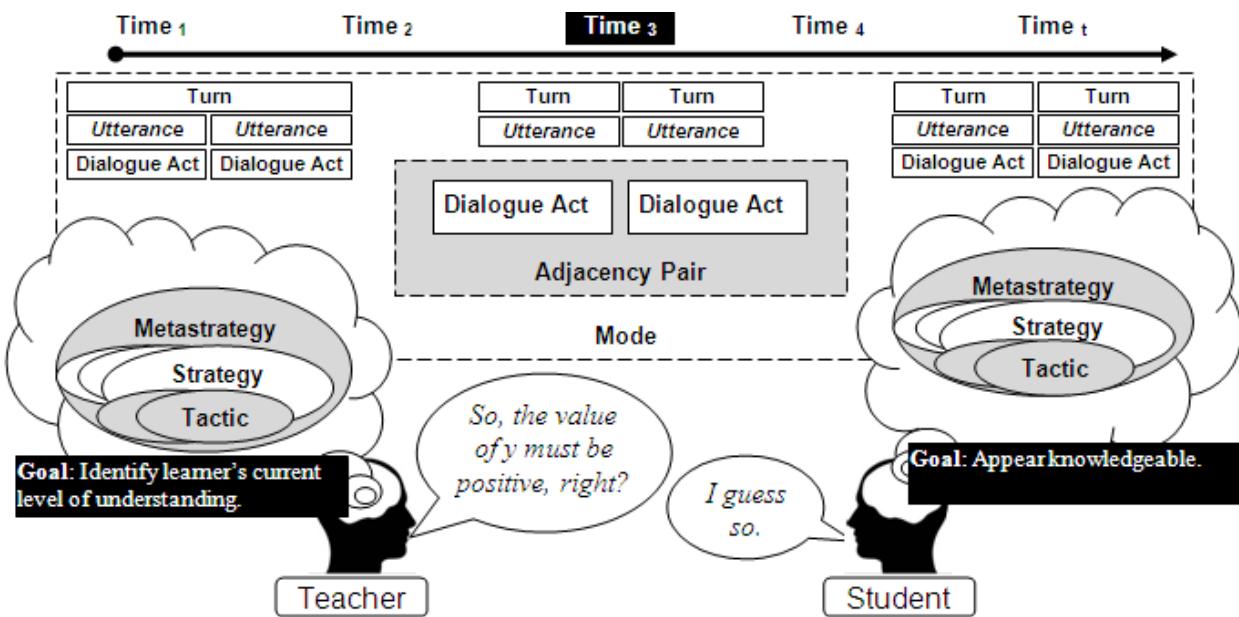
Figure 1. Anatomy of a tutorial dialogue.

Conversations also have higher-level structures, which, following Cade et al. (2008), we are calling *modes*. In addition to modes such as *openings* and *closings* (Schegloff, 1968), which are associated with human conversation in general, modes common to tutorial dialogues include *modeling*, *scaffolding*, and *fading*, representing various degrees of tutorial support. To repeat, a mode is not, in our way of thinking, a strategy. Rather, these instructional modes are, in a sense, "named after" strategies. They are identifiable sequences of utterances in which the tutor's moves (dialogue acts) may be interpreted as tactics aimed at carrying out a particular strategy, aimed at a particular goal.

So, within this framework, any given stretch of tutorial dialogue consists of a set of utterances, each of which may be classified as a distinct dialogue act. Certain dialogue acts are associated with certain modes, and some can trigger a switch from one mode into another. For example, the utterance "I think we're running out of time" can be understood as a dialogue act aimed at switching the conversation into closing mode. Also, as this example illustrates, dialogue acts are assumed to be intentional, and as such may be understood as *tactics* aimed at achieving certain goals. Further, it is assumed that any given tactic reflects a particular hidden strategy, selected from among a set of available strategies, in accordance with a metastrategy.

We are currently in the process of applying this framework to the analysis of a large dataset consisting of some 250,000 dialogue transcripts provided to us by Tutor.com, a leading provider of online, chat-based human tutoring. Using a coding scheme based on this framework, we will be training a panel of subject matter experts (SMEs), selected from among Tutor.com's most highly rated tutors, to hand-tag each utterance in a subset of approximately 1,200 transcripts. Each of the approximately 96,000 utterances will be coded as representing a particular dialogue act, and dialogue acts that are identified as mode switches will be coded as such. After reviewing for accuracy and internal consistency, we will use the resulting "gold standard" training set to train an automatic dialogue act classifier, which will be used to tag the remaining transcripts in the data set. We will then use sequencing and clustering algorithms to discover hidden patterns (interpretable as modes/strategies) associated with successful and less successful sessions – as established both by internal evidence of learning and the learner and tutor ratings available in the transcript metadata. In this way, we hope to identify the metastrategies that expert tutors use to help students learn.

Summary and Conclusion

In this chapter, we have proposed a conceptual framework for use in the analysis of any interaction between two humans where at least one of the participants seeks to gain, or impart, new knowledge or skill from or to the other. This framework builds from the ground up, beginning with the notion of a *move*, which we have defined as any action that has a bearing on the course of the interaction between participants. The analogy is to a move in chess, or a "throw" in the game Rock, Scissors, Paper. A *tactic* is a move with particular purpose, selected in a deliberate, principled way from a set of possible moves, in accordance with a *strategy*, which may be thought of as a more or less formal algorithm for choosing tactics based on changing circumstances. At the next level up, a *metastrategy* is an algorithm for selecting among strategies.

This framework, we suggest, is especially useful because it can be applied to the full range of human pedagogical interaction, including formal classroom instruction, informal tutoring, and the operations of computer-based learning systems such as intelligent tutors (both rule-based and dialogue-based) and serious games. Indeed, absent a common framework for describing the nature of the various instructional practices within these different kinds of systems, it is hard to imagine how our field will ever be able to

build the kind of "science-based instructional theory" that (Koedinger, Booth, and Klahr, 2013) call for. Indeed, a generalized architecture such as GIFT—which is aimed at supporting and integrating communication across systems as diverse as serious games, dialogue-based intelligent tutors, and other forms of adaptive learning systems—will especially benefit from, if not require, precise definitions for describing moves, tactics, strategies, and metastrategies, and how these are instantiated within and across systems.

References

Austin, J. L. (1965). How to do things with words. New York: Oxford University Press.

Cade, W. L., Copeland, J. L., Person, N. K. & D'Mello, S. K. (2008, January). Dialogue modes in expert tutoring. In Intelligent tutoring systems, (pp. 470-479). Springer Berlin Heidelberg.

Clark, H. H. & Schaefer, E. F. (1989). Contributing to discourse. Cognitive science, 13(2), 259-294.

Graesser, A. C., Chipman, P., Haynes, B. C. & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. IEEE Transactions on Education, 48(4), 612-618.

Graesser, A. C., Person, N. K. & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. Applied cognitive psychology, 9(6), 495-522.

Koedinger, K. R., Booth, J. L. & Klahr, D. (2013). Instructional complexity and the science to constrain it. Science, 342(6161), 935-937.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J. & Willingham, D. T. (2013). Improving Students' Learning With Effective Learning Techniques Promising Directions From Cognitive and Educational Psychology. Psychological Science in the Public Interest, 14(1), 4-58.

Olney, A. M., Person, N. K. & Graesser, A. C. (2012). Guru: Designing a Conversational Expert Intelligent Tutoring System. In C. Boonthum-Denecke, P. McCarthy & T. Lamkin (Eds.), Cross-Disciplinary Advances in Applied Natural Language Processing: Issues and Approaches (pp. 156-171). Hershey, PA: Information Science Publishing.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M. & Metcalfe, J. (2007). Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007-2004. National Center for Education Research.

Rus, V., Graesser, A., Moldovan, C. & Niraula, N. (2012). Automatic Discovery of Speech Act Categories in Educational Games, 5th International Conference on Educational Data Mining (EDM12), June 19-21, Chania, Greece.

Searle, J. R. (1969). Speech acts: An essay in the philosophy of language. Cambridge University Press.

Tomasello, M. (1999). The cultural origins of human cognition. Harvard University Press.

Vanlehn, K. (2006). The behavior of tutoring systems. International journal of artificial intelligence in education, 16(3), 227-265.

Vygotsky, L. (1978). Mind in society. Cambridge, MA: Harvard University Press. Original work published in 1930.

Wang, N., Johnson, W. L., Rizzo, P., Shaw, E. & Mayer, R. E. (2005, January). Experimental evaluation of polite interaction tactics for pedagogical agents. 10th international conference on Intelligent user interfaces (pp. 12-19). ACM.

CHAPTER 18 – Instructional Strategies in Trialogue-based Intelligent Tutoring Systems

Zhiqiang Cai, Shi Feng, Whitney Baer, and Arthur Graesser
University of Memphis

Introduction

The inception of conversational computer systems is often traced to ELIZA (Weizenbaum, 1966). ELIZA simulated a client-centered Rogerian psychotherapist that attempts to get the clients to talk about their problems and self-discover possible solutions. ELIZA syntactically converted the users' statements into questions and comments to encourage the users to express themselves. ELIZA had minimal linguistic and semantic resources, but rather relied on recognizing a finite class of keywords and a modest set of production rules to maintain the conversation. However, this simple mechanism surprisingly generated a reasonably smooth conversation. Shortly after that landmark system, several conversational systems were developed. The efforts were devoted to simulating specific agents (Colby, Weber & Hilf, 1971), tutoring specific domains (Collins, Warnock & Passafiume, 1975), and answering questions (Woods, 1977; Norman & Rumelhart, 1975).

Near the turn of the 21st century, learning environments started to incorporate animated conversational agents synchronized with high quality speech and appropriate gestures (Graesser, Lu, Jackson, Mitchell, Ventura, Olney & Louwerse, 2004; Johnson, Rickel & Lester, 2000). Dozens of such systems have been built during the last decade. For example, Graesser, Wiemer-Hastings, Wiemer-Hastings, Kreuz, and the Tutoring Research Group (TRG) (1999) created AutoTutor to teach computer literacy and physics; McNamara, Levenstein, and Boonthum (2004) developed iSTART to teach reading strategies; Biswas, Schwarts, Leelawong, Vye, and TAG-V (2005) built Betty's Brain to teach biology with teachable agents; and Halpern, Millis and Graesser (Millis, Forsyth, Butler, Wallace, Graesser & Halpern, 2011; Halpern, Millis, Graesser, Butler, Forsyth & Cai, 2012) developed Operation Acquiring Research Acumen (ARA) to teach critical scientific reasoning. All these systems use one or more agents in the conversations in an effort to help students learn.

Conversational agents in a learning environment can play different roles. When a system employs two conversational agents to interact with one human learner, the system provides a platform for three-party conversations. This kind of conversation is called a "trialogue" (Cai, Graesser, Forsyth, Burkett, Millis, Wallace, Halpern & Butler, 2011). The two computer agents are usually assigned the roles of a tutor and a peer student. The use of a tutor agent and a peer student agent enables different conversation modes. In this chapter, we discuss three trialogue modes that correspond to three instructional strategies: (1) vicarious learning, (2) expectation-misconception-tailored tutoring, and (3) learning by teaching a teachable agent.

Vicarious Learning

Vicarious learning is also called observational learning or modeling, which is a successful learning form, particularly for low knowledge learners (Bandura 1986, Craig, Sullins, Witherspoon & Gholson, 2006). In this form, students learn by observing other people's behavior. For example, a student learns by observing a teacher's problem solving process, a classmate's politeness in asking questions, a parent's house cleaning process, and so on. There is a substantial body of empirical research on vicarious learning. Bandura (1986) identified four stages in vicarious learning when the learner observes other human role

models: attention, retention/memory, initiation/motor, and motivation. Craig et al. (2006) were among the early researchers who incorporated vicarious learning into learning environments and investigated the consequences on comprehension and memory. Vicarious learning was successfully integrated into a conversational ITS, *Operation ARA* (Cai et al., 2011; Halpern et al., 2012). In Operation ARA, a student's knowledge about a certain topic is assessed through a set of multiple-choice questions. The vicarious learning mode is triggered when the assessment shows that the student does not have much knowledge about the topic.

Vicarious learning always involves a model. The model could be a teacher, a parent, or a peer student. In computerized learning environments, the model could be a tutor agent or a peer student agent. The tutor agent is often used as the model to show the process of solving a complex problem. The peer student agent can be used as a model to demonstrate a learning experience, including challenges and struggles. Tutor agent models can provide quick and direct instruction, which could be good for a quick learner. In contrast, peer student agent models may provide examples of learning experiences closer to the learners' models. Slower learners particularly may benefit from a peer student agent model, but there is no strong empirical support for this hypothesis. Peer student agent models may also have advantages in Bandura's four stages of vicarious learning, but empirical tests are needed to substantiate this notion.

In a conversational learning environment, a vicarious learning trialogue (VLT) can be meticulously designed to demonstrate a learning process involving a tutor agent and a peer student agent. In a VLT, the majority of the conversations transpire between the tutor agent and the peer student agent. The tutor agent gives the peer student agent instructions, explanations, hints, answers to the peer student agent's questions, and corrections of the misconceptions that the peer student demonstrates. The peer student agent demonstrates how a typical learner may acquire the knowledge, including difficulties. The peer student agent asks questions, responds to hints, sometimes improves answers, and sometimes exhibits errors. Eventually, the two agents provide the correct solution to the problem.

The VLT strategy is typically used to support a learner with a low-level of knowledge. However, it is also important to involve the learner by giving them the opportunity to answer some easy questions or provide solutions to some easy parts of the problem. The purpose of doing this is to make sure the learners' minds do not wander and that they engage in the conversation.

We illustrate a VLT with the following conversation involving Lisa (a tutor agent), Tom (a peer student agent), and David (a human learner):

A01: **Lisa**: In the video we watched just now, a packet was released from an airplane when it was right above the target. Do you think the packet will hit the target, Tom?

A02: **Tom**: Yes, I do.

A03: **Lisa**: Oh, really? Why do you think so?

A04: **Tom**: After the packet is released, there will be no more force acting on the packet, so the packet will drop vertically down to the ground and thus hit the target.

A05: **Lisa**: Why should the packet drop down?

A06: **Tom**: Because the gravity will pull it down. Oh, yes, I realized that there is force acting on the packet after it is released.

A07: **Lisa**: What is that force?

A08: **Tom**: Gravity.

A09: **Lisa**: Wonderful! After release, gravity is the only force acting on the packet.

A10: **Tom**: And that is why the packet will drop directly down.

A11: **Lisa**: David, do you agree with Tom?

A12: **David**: What?

A13: **Lisa**: Tom thinks that the packet will hit the target. Do you think so, too?

A14: **David**: It sounds right.

A15: **Lisa**: Hmmmm. Let's think more about this. Before the packet was released, did the packet have a horizontal velocity, Tom?

A16: **Tom**: Oh, yes, it had the same horizontal velocity as the plane.

A17: **Lisa**: Right. To change the horizontal velocity, the force acting on the packet must have a non-zero horizontal component. Does the gravity have a non-zero horizontal component?

A18: **Tom**: No. The horizontal component of the gravity is zero.

A19: **Lisa**: That is right. Then will the packet's horizontal velocity change?

A20: **Tom**: No. It will keep the same horizontal velocity when it is dropping down.

A21: **Lisa**: Agree. Then where will the packet be when it touches the ground?

A22: **Tom**: It will travel some horizontal distance and miss the target.

A23: **Lisa**: David, do you agree?

A24: **David**: Yes, I agree.

A25: **Lisa**: Good thinking. I have the impression you both understand this pretty well.

In the turn *A01*, Lisa, the tutor agent, initiates the conversation with a question. Then, Tom, the student agent, gives a typical incorrect answer that many learners might give. Lisa then asks a "Why" question to Tom. Tom's explanation in *A04* shows a typical misconception. Lisa gives a hint in *A05* and Tom corrects the misconception in *A06*. *A07* to *A10* further clarifies and confirms a partial answer to the solution. The human learner is involved in *A11* to *A14*. Notice that, the "What" response may indicate that the human learner's mind was wandering. Therefore, *A11* to *A14* may help the learner engage in the learning process. The user's attention is directed to the conversation by the use of the human learner's name. The rest of the conversation covers another aspect of the problem using a similar conversation pattern. The whole conversation demonstrates to the learner how a typical student might figure out an answer with the help of a tutor. The above conversation was simplified to save space here. In the real system, the conversation is much longer in order to cover all expected answers in the problem.

VLT is the least challenging conversation mode for learners. Strategically, we think it should be used only when a learner has a low-level of knowledge about the intended subject. In Operation ARA, the learner's

knowledge level about the topic would have been assessed before the VLT mode is triggered. A higher level assessment module is needed to determine the leaner's knowledge level about the topic. In turn, VLT can further assess the student's knowledge during the conversation. However, VLT is usually weak in assessing a learner's deep comprehension because learners only have minimal interactions in the conversation.

VLT is also the least challenging kind of conversation for script authors to design. VLT scripts can be written in a linear form with some specific locations where the learner's input is expected. For each input location, a few alternative tutor contributions need to be prepared to handle different types of input. For example, if in *A24*, David had said "Disagree," Lisa would respond in *A25* with something like "Well, actually, Tom was right. The packet will miss the target." VLT can also be designed so that it provides many versions of the same conversational context. With such a design, the learner may have a new experience when the learner revisits the topic. Developing VLT is relatively easy because the learners' answers are usually easy to parse. However, computational linguistic support is still helpful. For example, David may say "Agree" or "Disagree" in many different ways. Accurately identifying the alternative wordings that are correct requires linguistic support.

While VLT is designed for low knowledge learners, it does not mean that the content in the conversation between the computer agents should only target shallow knowledge. Instead, VLT can be a better way to deliver deep knowledge to low knowledge learners. Craig et al. (2006) has provided evidence that vicarious presentation is generally effective when it is used in deep-level reasoning questions. More evidence about the effect of VLT can be expected through future analysis of data from Operation ARA, GIFT, and other systems that use VLT.

Expectation and Misconception Tailored Tutoring

Expectation and misconception tailored (EMT) conversation is an effective conversation strategy for intelligent tutoring (Graesser et al., 1999, Graesser, D'Mello, Hu, Cai, Olney & Morgan., 2012). An EMT conversation targets a complex problem. The ideal answer to such a problem can be split into multiple parts, called "expectations." For example, the equation *Force equals Mass times Acceleration* involves the learner understanding that the force involves both mass and acceleration. Therefore, the learner must correctly answer both parts when asked the question: "What is Force equal to?" The goal of the conversation is to help learners to construct an answer that covers all expectations. Correcting learners' misconceptions about the problem is another goal. During the conversation, the system keeps checking the coverage of expectations and detecting any of the learner's misconceptions by analyzing the learner's verbal contributions.

In dialogue-based systems, an EMT conversation is between a tutor agent and a human learner. The tutor agent initiates a conversation by asking a main question. Then the learner and tutor collaboratively answer the question through multiple conversational turns. The system evaluates the learner's contributions over each of the conversational turns. The tutor agent also gives hints and prompts to help the learner improve the answer. Just like ELIZA, the tutor agent tries to get the student to do the talking.

In a trialogue system with the addition of a peer student agent, the conversation can be more engaging, if not entertaining. One problem that has challenged researchers is distinguishing expectations and misconceptions. The verbal form of a misconception is sometimes very close to an expectation, which makes it very difficult to differentiate. A semantic algorithm may identify a misconception only to a certain degree. Consequently, any time the system decides to correct a misconception, it could be the wrong decision. This creates a dilemma. On a personal level, it may be irritating to a human when the learner sees his/her own good answer treated as a misconception. This problem is solved neatly by trialogues.

When a misconception is detected by the system, the peer student agent chimes in and articulates a system-stored misconception that is semantically close to the human's input. The tutor then critiques the peer agent's misconception, instead of the human's. The student agent takes the blame for a misconception rather than the human.

The peer student may also contribute alternative good answers, entertaining comments, and speech acts to redirect the conversation. The following simplified conversation demonstrates an EMT trialogue among Lisa (a tutor agent), Tom (a peer student agent), and David (a human learner):

*B01***: Lisa:** In the video we just watched, a packet was released from an airplane when it was right above a target. Do you think the packet will hit the target, David?

*B02***: David:** Yes, I do.

*B03***: Lisa:** Oh, really? Why do you think so?

*B04***: David:** Because there is no force on the packet.

*B05***: Lisa:** What do you think, Tom?

*B06***: Tom:** I agree with David. After the packet is released, there will be no force acting on the packet, so the packet will drop vertically down to the ground and hit the target.

*B07***: Lisa:** I'm not sure that's right, Tom.

*B08***: Lisa:** What do you think, David?

*B09***: Lisa:** Why should the packet drop down?

*B10***: David:** Gravity pulls the packet down.

*B11***: Lisa:** That is correct, David.

*B12***: Lisa:** Tom, what do you think about that?

*B13***: Tom:** Oh, yes, I realized that there is force acting on the packet after it is released.

*B14***: Lisa:** What force is that?

*B15***: Tom:** Gravity.

*B16***: Lisa:** Exactly.

*B17***: Lisa:** After release, gravity is the only force acting on the packet.

B18: **Lisa:** OK. Let's think more about this.

B19: **Lisa:** Before the packet was released, did the packet have a horizontal velocity, David?

*B20***: David:** Yes. Same as the plane.

B21: **Tom:** Oh, yes, I agree. The packet had the same horizontal velocity as the plane before it was released.

B22: **Lisa:** Right.

B23: **Lisa:** To change the horizontal velocity, the force acting on the packet must have a non-zero horizontal component. Does the gravity have a non-zero horizontal component, David?

B24: **David:** No. It is zero.

B25: **Tom:** I think so, too. The horizontal component of the gravity is zero.

B26: **Lisa:** Right.

B27: **Lisa:** Think about this.

B28: **Lisa:** Will the packet's horizontal velocity change after release, David?

B29: **David:** No. Probably not. I am not sure.

B30: **Tom:** You look hesitant, David. I would say, no! It won't change.

B31: **Lisa:** I agree.

B32: Lisa: The packet's horizontal velocity will stay the same as the plane after release.

B33: Lisa: There is still more to think about.

B34: Lisa: Where will the packet be when it hits the ground, David?

B35: David: Somewhere away from the target.

B36: **Tom:** Yeah, it will miss the target.

B37: **Lisa:** Right.

B38: **Lisa:** It will travel some horizontal distance and miss the target when it touches the ground.

B39: **Lisa:** Let's continue the movie and see what happens.

In the turn *B01*, Lisa initiates the conversation with a "yes/no" question and follows with a "why" question in *B03*. In *B04*, David's answer contains a misconception. Tom (peer student agent) restates the misconception in a more complete form in *B06*. Lisa gives negative feedback in *B07* and starts the misconception correction process. The misconception is fully corrected in *B17*, which also covers an expectation. The turn *B18* is a conversation advancer that directs the conversation to another aspect of the problem. The answer of the learner in *B20* is correct. Tom restates the answer in *B21* and makes it a complete answer. Lisa gives positive feedback in *B22*. *B23*, *B28* and *B34*, pointing to different expectations. The conversation ends with all expectations covered.

In EMT trialogues, the learner is deeply involved in the conversation. The learner is responsible for constructing an acceptable answer to the question. Therefore, EMT is more challenging to learners than

VLT. As a strategy, we hypothesize that this trialogue mode should be used when the learner has a moderate-level of knowledge about the intended subject.

EMT scripts are more complex than VLT scripts. An EMT script usually contains a main question and multiple expectations, misconceptions, hints, prompt questions to elicit specific words, and various types of answers to each question. Script authors start by composing a main question, followed by an ideal answer. The ideal answer is then split into expectations, each of which contains some hints and prompts with possible good answers and bad answers. A list of misconceptions is also identified and corresponding corrections are composed. More details about the script structure can be found in Graesser et al. (2012).

EMT is more challenging for software developers. The system needs to have enough computational linguistics resources to support an EMT conversation. For example, a speech act classifier and a semantic analysis engine are necessary. Authoring such scripts requires well-organized authoring tools, as in the case of the AutoTutor Script Authoring Tool (ASAT) (Cai, Hu & Graesser, 2013).

The effect of EMT mode on learning has been replicated in many evaluations of AutoTutor systems. The effect size for single agent one-on-one tutoring is 0.8 (Graesser et al., 2012). The effect size of EMT in a trialog mode shows a comparable effect size (Forsyth et al., 2012), but needs more systematic replications.

Learning By Teaching A Teachable Agent

The term "teachable agent" has been used to refer to intelligent computer agents that can be "taught" by the human to perform some tasks or solve some problems (Biswas et al., 2005). The idea is to provide opportunities for students to learn by teaching. In Operation ARA, the term "teachable agent" was used to refer to a specific trialogue mode (Cai et al., 2011; Halpern et al., 2012). In teachable agent trialogue (TAT) mode, the peer student agent takes on the role of a teachable agent and continually seeks help from the human learner. The tutor agent provides assistance when the human learner has difficulty. The majority of the conversation is between the peer student and the human learner.

The following conversation illustrates a TAT among Lisa (tutor agent), Tom (peer student agent), and David (human learner):

C01: **Lisa:** In the video we just watched, a packet was released from an airplane when it was right above a target. Do you think the packet will hit the target, Tom?

C02: **Tom:** I have no idea.

C03: **Tom:** David, can you help me on that?

C04: **David:** The packet will miss the target.

C05: **Tom:** How come, David? Wouldn't the packet drop straight down?

C06: **David:** Of course not.

C07: **Tom:** Really? Why, David?

C08: **David:** Because the packet has a non-zero horizontal velocity.

C09: **Tom:** Before release or after release?

C10: **David:** The same horizontal velocity before and after release.

C11: **Tom:** I can't imagine that. Is that true, Lisa?

C12: **Lisa:** It is true. After the packet is released, the packet will keep a constant horizontal velocity same as the plane.

C13: **Tom:** Oh, but why? Is there any force acting on the packet other than the gravity?

C14: **Lisa:** I will let David explain that to you.

C15: **David:** Gravity is the only force and the horizontal component is zero. That is why the horizontal velocity stays constant.

C16: **Tom:** I see. I remember that is one of Newton's laws.

C17: **Lisa:** Newton's first law of motion.

C18: **Tom:** Oh, I got it. Thank you, David.

C19: **David:** You are welcome!

C20: **Lisa:** I enjoyed talking with you guys. This is hard material, but you seem to understand it well.

In the above simplified conversation, Lisa initiates the question in *C01*. Tom turns to David for help in *C03*. Tom keeps asking David for deeper explanations of the problem. In *C11*, Tom turns to Lisa to give Lisa a chance to present a more accurate explanation. In *C13*, Tom shifts to another aspect of the problem and, in C14, Lisa redirects the conversation to David. In *C17*, Lisa provides important assistance. The conversation could be much longer in a real system.

In TAT mode, the learner's role is a "teacher." Therefore, the learner needs enough knowledge about the subject and skills in communication in order to make the conversation successful. Because of that, this strategy should be used for learners with a high-level of knowledge.

The script authoring process for TAT can be very similar to EMT. The difference is that the hints and prompts need to be prepared for the peer student agent. Therefore, the questions should sound like they came from a leaner, instead of a teacher.

Like EMT, TAT also needs the support of high performance computational linguistics modules. The system needs to be able to assess covered and uncovered aspects of the problem. For example, before the peer student agent asks the question in *C09*, the system needs to determine that the learner's last utterance was about the horizontal velocity, the learner already mentioned that the horizontal velocity is non-zero, and the learner has not yet mentioned "before release" or "after release". In *C13*, there is a "but why" before the peer student agent's question. To use that correctly, the system needs to determine that the learner's last utterance contains a statement that needs further explanation and that a good answer can be provided to the "but why" question.

The effectiveness of teachable agents on learning has been substantiated in systems such as *Betty's Brain* (Biswas et al., 2005). Forsyth et al. (2012) and Halpern et al. (2012) report positive effects for Operation ARA, but more research is needed in the future.

Discussion and Recommendations

We have discussed the conditions for strategic use of the three basic trialogue modes, but the effect of each mode on student learning is not completely settled empirically. An adequate evaluation of these trialogue modes would require an assessment of the trialogue mode selection mechanism, an assessment of the quality of trialogue implementation, and an assessment of the learner's knowledge and their learning gains. Systems with high quality implementation of trialogues need to be developed prior to any research experiments.

Other trialogue modes are possible. For example, Operation ARA has a competition mode that is set up between the peer student agent and the human learner (Halpern et al., 2012). The tutor agent asks a question, then the peer student agent and the human learner each gives an answer. The tutor judges the answers and gives a correct answer. This mode is expected to be more effective in engaging learners. As another example, trialogues have been developed to stage disagreements between the tutor agent and student agent, with the hopes that the human learner can help sort out the correct answer and explain why it is correct (D'Mello, Lehman, Pekrun & Graesser, in press; Lehman, D'Mello, Strain, Mills, Gross, Dobbins, Wallace, Millis & Graesser, 2013). The disagreement puts the learner in cognitive disequilibrium, which often triggers confusion, reasoning to resolve the disagreement, and deep learning.

AutoTutor provides a unified way (Cai et al., 2011) to implement all of the trialogue modes presented in this chapter. To facilitate the use of AutoTutor, ASAT integrates various functions such as script creation, conversation rule configuration, script validation, and automatic tutoring session simulation. ASAT provides a way to set up trialogue conversations that can be driven by user's verbal input as well as any other user activities, such as choosing an answer to a multiple-choice question, dragging and dropping an object in a page, etc. However, setting up conversation rules is still complex and difficult for non-programmer users. Additionally, AutoTutor Web Service (ATWS) provides a publicly accessible engine for script interpretation (Cai et al., 2013). AutoTutor script can be remotely submitted to ATWS for continuous interpretation. Both ASAT and ATWS are accessible through GIFT.

Trialogue strategies can be extended to environments involving more than two computer agents. However, conversations involving three or more human learners are beyond the scope of this chapter. As we have pointed out, trialogue-based conversations have been used in tutoring and assessment systems. ASAT and ATWS make it possible for GIFT users to use trialog conversations in their own systems. When a trialogue conversation is used, it is important to remember the use of the strategies presented in this chapter. GIFT should provide a persistent student model to guide conversation strategy selection. GIFT users may develop new trialogue strategies or more adaptively use the strategies we presented in this chapter, but it is important to assess the effects of each strategy for readers with different abilities. GIFT provides a perfect platform for strategy assignment and assessment.

Acknowledgments

References

Bandura, A. (1986). Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall.

Biswas, G., Schwarts, D., Leelawong, K., Vye, N. & TAG-V. (2005). Learning by teaching: A new agent paradigm for educational software. Applied Artificial Intelligence, 19, 363-392.

Cai, Z., Graesser, A. C., Forsyth, C., Burkett, C., Millis, K., Wallace, P., Halpern, D. & Butler, H. (2011). Trialog in ARIES: User input assessment in an intelligent tutoring system. In W. Chen & S. Li (Eds.), Proceedings of the 3rd IEEE International Conference on Intelligent Computing and Intelligent Systems (pp.429-433). Guangzhou: IEEE Press.

Cai, Z., Hu, X. & Graesser, A. C. (2013). ASAT: AutoTutor script authoring tool. Paper presented at the meeting of the Society for Computers in Psychology, Toronto, CA.

Colby, K. M., Weber, S. & Hilf, F. D. (1971). Artificial paranoia. Artificial Intelligence, 2, 1-25.

Collins, A., Warnock, E. H. & Passafiume, J. J. (1975). Analysis and synthesis of tutorial dialogues. In G. H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory (Vol. 8, pp. 49-87). New York: Academic Press.

Craig, S. D., Sullins, J., Witherspoon, A. & Gholson, B. (2006). Deep-level reasoning questions effect: The role of dialog and deep-level reasoning questions during vicarious learning. Cognition and Instruction, 24(4), 565-591.

D'Mello, S. K., Lehman, B., Pekrun, R. & Graesser, A.C. (in press). Confusion can be beneficial for learning. Learning and Instruction.

Forsyth, C.M. Pavlik, P., Graesser, A.C. Cai, Z., Germany, M., Millis, K., Butler, H., and Dolan, R. (2012). Learning gains for core concepts in a serious game on scientific reasoning. In K.Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson & J. Stamper (Eds.) Proceedings of the 5th International Conference on Educational Data Mining (pp 172-175). Chania, Greece: International Educational Data Mining Society.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R. & the Tutoring Research Group. (1999). AutoTutor: A simulation of a human tutor. Journal of Cognitive System Research, 1, 35-51.

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A. M. & Louwerse M. M. (2004). AutoTutor: A tutor with dialogue in natural language. Behavior Research Methods, Instruments & Computers, 36, 180-193.

Graesser, A. C., D'Mello, S.K., Hu, X., Cai, Z., Olney, A. M. & Morgan, B. (2012). AutoTutor. In P. M. McCarthy & C. Boonthum (Eds.), Applied natural language processing and content analysis: Identification, investigation and resolution (pp. 169-187). Hershey, PA: IGI Global.

Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C. & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. Thinking Skills and Creativity, 7, 93-100.

Johnson, W. L., Rickel, J. W. & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. International Journal of Artificial Intelligence in Education, 11, 47-78.

Lehman, B., D'Mello, S. K., Strain, A., Mills, C., Gross, M., Dobbins, A., Wallace, P., Millis, K. & Graesser, A. C. (2013). Inducing and tracking confusion with contradictions during complex learning. International Journal of Artificial Intelligence in Education, 22, 85-105.

McNamara, D. S., Levenstein, I. Bl & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. Behavioral Research Methods, Instruments, and Computers, 36, 222-233.

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A. C. & Halpern, D. (2011) Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou & J. Lakhmi (Eds.) Serious games and edutainment applications (pp.169-196). London, UK: Springer-Verlag.

Norman, D. A. & Rumelhart, D. E. (1975). Explorations in cognition. San Francisco: Freeman.

Weizenbaum, J. (1966). ELIZA – A computer program for the study of natural language communication between man and machine. Communications of the ACM, 9, 36-45.

Woods, W. A. (1977). Lunar rocks in natural English: Explorations in natural language question answering. In A. Zampoli (Ed.), Linguistic structure processing (pp. 201-222). New York: Elsevier.

CHAPTER 19 – Where in the Data Stream Are We?: Analyzing the Flow of Text in Dialogue-Based Systems for Learning

Donald M. Morrison, Benjamin Nye, and Xiangen Hu
Institute for Intelligent Systems, University of Memphis

Abstract

This chapter continues a discussion we began in the first volume in this series (Hu, Morrison & Cai, 2013) concerning the use of "learner micromodels" in dialogue-based ITSs. As we originally defined it, a learner micromodel in an ITS is an estimate of a learner's cognitive and/or affective state at a given time in an ITS session, based entirely on the real-time dynamics of that particular session. An example of such a micromodel, which we call a Learner's Characteristic Curve (LCC), tracks two features of a learner's recent dialogue history – *novelty* and *relevance* – where novelty is a measure of the degree to which the learner's contributions to the dialogue add something new, and relevance is a measure of the degree to which his or her contributions conform to an expected answer. This model is called a "characteristic curve" because the learner's trajectory on these measures over time can be evaluated against certain archetypal curves or boundaries (e.g., thresholds). In this chapter, we identify an interesting complication with this design, then expand the discussion to include some general principles concerning the analysis of the data sets and live streams of data that are beginning to flow in vast quantities from Internet-based learning environments, including those with human tutors, artificially intelligent tutors, and, perhaps most interestingly, hybrid systems of the future.

Introduction

In Hu, Morrison & Cai (2013), we described a method for tracking a learner's contributions to a tutorial dialogue on two dimensions: the *relevance* of the contribution (calculated as the degree of semantic similarity to an expected answer) and its *novelty* (semantic similarity to the learner's previous contributions). These LCC dimensions can be used by an ITS as inputs to algorithms that make instructional choices. The LCC approach is particularly well-suited for a multiagent ITS design (Nye & Morrison, 2013). For example, assume a multiagent tutoring system with an LCC calculation agent, a learner modeling agent, and a tutoring agent. The LCC agent can calculate LCC scores for a learner and broadcast them to other agents. The learner modeling agent can track the LCC scores for a given learner over time, then use this information to build, and continuously update, a learner profile for some longer-term purpose. A tutoring agent can employ the LCC information to help select its next dialogue move. If both relevance and novelty scores remain low (i.e., the learner repeatedly gives the same incorrect answer), the agent can be programmed to provide the expected answer, while if relevance is low but novelty is high (the learner keeps trying), the agent can decide to give a prompt or a hint. Importantly, the LCC agent does not need to coordinate with other agents or even "know" how its outputs are used. As such, LCCs make suitable ITS components for use in relatively lightweight, loosely coupled modules. A modular ITS can combine small, partial solutions that become progressively more effective, breaking into smaller, more tractable pieces the notoriously difficult problem of intelligent natural-language tutoring.

In this chapter, we pursue these ideas further, using the LCC construct to leverage a discussion of larger issues related to the development of intelligent systems for learning, including hybrid systems involving

both human tutors and artificially intelligent ones. We particularly focus on the packets of data that will flow through the communication channels of next-generation, multiagent learning systems that support tutorial dialogue. We define these ITSs as including those that use natural-language conversation between two agents, at least one of which has a pedagogical intent, and at least one of which is human.[1] We begin by reviewing the workings of the LCC agent as originally defined, then identify a key limitation for this design: the need to determine if inputs to a tutorial dialogue are true "contributions" (conceptual explanations) or something else. This leads into a more general discussion concerning the analysis of data streams generated by online pedagogical conversations and the potential for data-mining those streams to enhance intelligent tutoring systems, online human-to-human tutoring, and hybrid systems that combine both approaches.

Learners' Characteristic Curves: Tracking Relevance and Novelty

The LCC technique was originally developed for use with AutoTutor Lite (Hu, Cai, Han, Craig, Wang & Graesser, 2009), a dialogue-based ITS developed at the Institute for Intelligent Systems. AutoTutor Lite is a simplified version of AutoTutor, a system designed to scaffold students' "deep" explanations of concepts (Graesser, Olney, Haynes & Chipman, 2005). This approach is partly based on research findings suggesting that students who *self-explain* are more likely to master the concept than those who don't (e.g., Chi, Lewis, Reimann & Glaser, 1989; VanLehn, Jones & Chi, 1992). While AutoTutor uses a number of additional strategies, AutoTutor Lite focuses primarily on self-explanation. In a typical use case, learners read a textbook-style presentation of a concept (such as Newton's first law of motion), after which AutoTutor Lite engages them in a dialogue aimed at eliciting their understanding of the concept in their own words. A complete and accurate summary of the concept, as stored by the intelligent tutor, is known as the *expectation*, and the utterances elicited from the learner are called *contributions*. The system relies on Latent Semantic Analysis (LSA) (Dumais et al., 1988; Graesser, Hu, Person, Jackson & Toth, 2002), which is used to calculate the semantic similarity between two pieces of text. Identical texts have a similarity score of 1, while non-overlapping texts (no shared meaning) have a score of approximately 0. In this way, a given learner contribution can be evaluated against the ideal student answer, giving a measure of *relevance* for each contribution. Using the same technique, each contribution can be compared to the learner's previous contributions, thus giving a measure of *novelty*.

As shown in the example in Table 1, combining these two measures gives four scores for a given utterance.

Table 1. Example LCC Scores for a Student Contribution.

	Old	**New**
Relevant	0.4 (O-R)	0.2 (N-R)
Irrelevant	0.1 (O-IR)	0.3 (N-IR)

Combining the relevance and novelty scores for the most recent utterance scores (O-R + N-R) gives a current relevant contribution score (CRC), while an overall *total coverage* score based on all the learners' contributions to the question so far. When calculated over a series of contributions, the scores on all six dimensions constitute what we have called a LCC.

---

[1] We do not mean to deny the plausibility of natural-language conversation only among artificially intelligent agents, nor the possibility that at some point such conversations might benefit systems that, to our knowledge, have not yet been created.

Importantly, the LCC model, as defined here, is in fact only one possible set of "learner characteristic curves." For example, additional time-series curves could be used to capture components of a learner's affect, as measured by dimensions such as boredom and confusion. The key idea is this: In their ongoing attempts to make sense of each other's utterances, human interlocutors do not just process each other's most recent move in isolation; rather, moves are analyzed in the context of previous moves, which must be kept in memory. It is in this way that interlocutors are able to maintain a sense of topic, and detect changes of topic, inconsistencies in assertions, needless or intentional repetition, and so forth. More generally, interlocutors need the history of the dialogue to build and update a theory about what is going on in the other's head, (i.e., the hidden "mental state"). In dialogues with a pedagogical purpose, where one of the interlocutors is a learner, this theory about what the learner is currently thinking and feeling may be referred to as a *learner micromodel*. The term *micromodel* is intended to distinguish it from a full learner model, which may be thought of as a theory about the whole learner, which a teacher or ITS develops over the course of a series of interactions. To gain an overall sense of the learner over time, information contained in the micromodel must eventually be incorporated into the long-term learner model, somewhat analogous to how the contents of working memory get transferred to longer-term storage.

Figure 1 illustrates the immediate context of the LCC technique. Each of the learner's moves (contributions) is evaluated on the basis of its relevance to the expectation (here, a full description of the features of parabolas) and its novelty in respect to previous contributions – as set by the boundaries of the LCC moving window, which determines the length of the history of utterances that are included in the analysis.
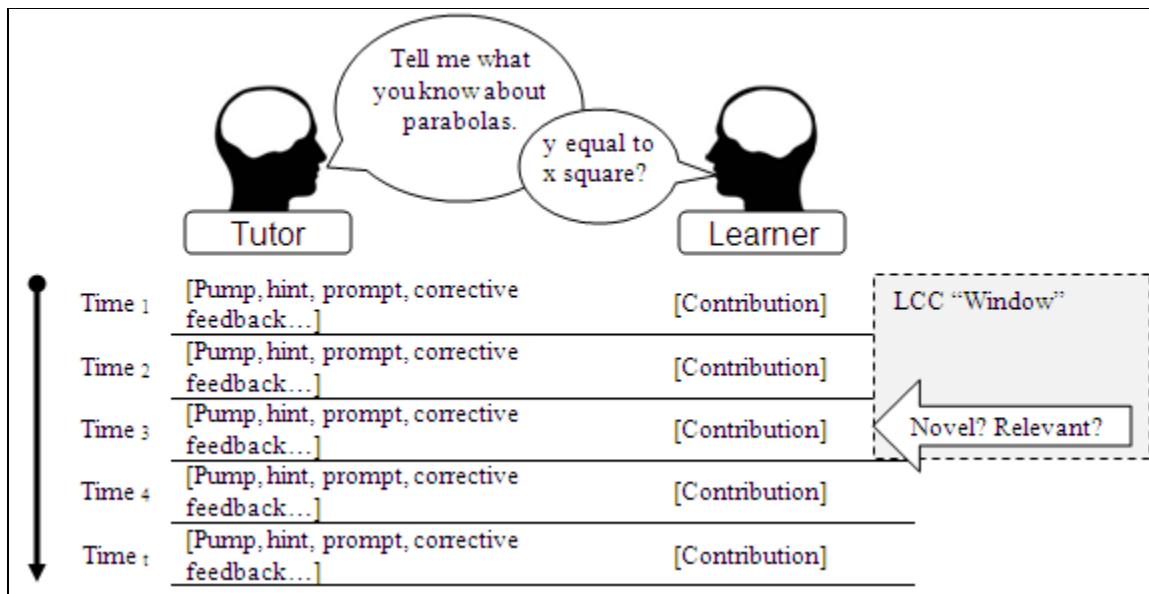
Figure 1. LCC moving window.

While the technique works in theory, there is an interesting practical complication. Put simply, a measure of novelty and relevance makes sense only if a given learner utterance is intended to contribute to an explanation of the concept under discussion. However, unless the learner is only allowed to make such assertions, some of the learner's utterances will be intrinsically irrelevant because they are not intended to answer questions or explain concepts. Moreover, if the focus of attention shifts, then the back edge of the window (i.e., the pointer to the first utterance in the dialogue history under scrutiny) needs to be reset. To illustrate the problem, consider the hypothetical dialogue transcript given below, where each utterance is labeled with a hypothetical dialogue act tag, implying a coding scheme.

1 **Tutor:**: You understand what a parabola is, right?
2 **Learner**: Right.
3 **Tutor**: Good.
4 **Tutor**: Tell me what you know about parabolas.
5 **Learner**: y equal to x square?
6 **Tutor**: Are you saying that the graph of the equation y=x^2 is a parabola?
7 **Learner**: yes?
8 **Tutor**: Great!
9**Tutor**: What else?
10 **Learner**: Say something else about parabolas?
10 **Tutor**: Sure
11 **Learner**: Okay
12 **Learner**: Bell-shaped?
13 **Tutor**: Are you saying that a parabola is shaped like a bell?
14 **Learner**: Yes
15 **Tutor**: So, why is a parabola shaped like a bell?
16 **Learner**: Not sure.
17 **Tutor**: Want to talk about that?
18 **Learner**: Okay
19 **Tutor**: Good.
20 **Tutor**: Here's a table showing some possible values of x and y in the case that y=x^2.
21 **Tutor**: Take a look and tell me what you see.

As can be seen, the learner makes two important contributions over the course of this dialogue that are both *relevant* and *novel*: parabolas have something to do with the equation *y=x^2* (line 5) and parabolas are bell-shaped (line 12). Presumably this information ought to be recorded in the tutor's developing theory of what the learner currently knows, i.e., the micromodel. However, these two contributions are separated in the dialogue history by utterances that clearly have other purposes. First, the learner tentatively agrees with the tutor's restatement of the learner's contribution (line 7), and second, at line 10, the learner checks her understanding of the tutor's prompt (*"What else?"*) by restating it (*"Say something else about parabolas?"*). It is only after this important business is gotten out of the way that the learner makes the second relevant and novel contribution. Because the intervening dialogue acts are not *assertions* (or *guesses*) about the concept under discussion, it does not make sense, and would be misleading, to evaluate them for relevance and novelty. In other words, the validity and accuracy of the novelty-relevance measure can be improved by preceding it with an algorithm for identifying and tagging utterances as representing specific dialogue acts, i.e., a *dialogue act classifier*.

Note also that near the very end of the transcript, around line 20, there is a distinct shift in focus, marked by the introduction of a new attentional scene: the table of possible values of *x* and *y*. Arguably, this is a point in the dialogue where it makes sense to reset the LCC, so that it begins monitoring the learner's contributions in respect to this new attentional scene, decoupling it, in some way, from the preceding segment. While an ITS can often handle this by dominating the selection of conversational topics (i.e., giving the user no choice), this is not always ideal. Moreover, for human-to-human dialogues, there is no "cheat sheet" of dialogue topics other than whatever exists in the participants' minds. In other words, while a dialogue act classifier can help to improve the validity of the LCC, it cannot alone help to decide when a particular segment in a dialogue has come to a close and another begun, as in the case of the changed attentional scene described above. What is needed is some way to track patterned sequences of dialogue acts, possibly sequences such as those that have been identified in the literature as *dialogue*

modes (D'Mello, Olney & Person, 2010; Cade, Copeland, Person & D'Mello, 2008). Finally, it is important to point out that a novelty-relevance measure only makes sense in the context of a tutorial strategy that is focused on eliciting an explanation or description of a particular concept. If an ITS employs this strategy alone (as does AutoTutor Lite), then an LCC based on measures of novelty and relevance makes good sense. But what if the tutor changes strategies, such that instead of eliciting a summary of a concept it simply walks the learner through an explanation of the concept one feature at a time, with the user only answering short verification (Yes/No) questions or checks for understanding? Dialogues in this form, such as interactive lectures, are common among expert tutors (D'Mello et al., 2010). In this case, it is the tutor who contributes most of the explanations and so the novelty and relevance of the contributions are less useful indicators.

Sequencing Tutorial Dialogues in Large Databases

Now let's step up a level and think about the larger context of conversations such as in the example above, the one regarding the nature of a parabola. Let's suppose, first of all, that the conversation in our example took place as an online chat session, which was recorded electronically and saved in a database along with transcripts of other such conversations. Second, let's suppose the tutor was *either* a human or an intelligent computer program.

Both cases are important. Let's assume first that the transcript represents a dialogue between a learner and human tutor. Much has already been learned from processing dialogue corpora consisting of some hundreds of such transcripts (e.g., D'Mello et al., 2010; Forbes-Riley & Litman, 2008; Graesser & Person, 1994; Graesser, Person & Magliano, 1995). The majority of such work has relied on manually coding the dialogue acts and content of such transcripts. This is efficient and effective with relatively small corpora. But what if there are *millions* of transcripts?

The emergence of the Internet as a worldwide medium of electronic communication has implications for the field of intelligent tutoring that are only beginning to become clear, or even felt. The pedagogically relevant online conversations taking place all around the world every minute of every day are producing massive quantities of text, much of which ends up in electronic storage and is potentially available for analysis. For example, Tutor.com, one of the world's leading providers of online human tutoring, employs some 3,000 tutors who have, in aggregate, held more than 10 million sessions, lasting an average of approximately 20 minutes each (Miller, 2013). Analysis of a database of this magnitude can potentially lead to new levels of understanding about how humans learn (or fail to learn) from each other, if only we can get machines to process very large quantities of natural language conversation, then report back to us in ways that make sense.

In fact, over the past 10 or 15 years, researchers with an interest in tutorial dialogue and ITSs have begun to employ machine learning techniques to extract knowledge from reasonably large dialogue-oriented text corpora (Boyer et al. 2009; Boyer et al., 2011; Chi, VanLehn, Litman & Jordan, 2011; Litman, Moore, Dzikovska,& Farrow, 2010; Olney et al., 2003; Rasor, Olney & D'Mello, 2011; Rosé et al., 2008; Rus et al., 2011; Rus et al., 2012). All involve some combination of human tagging and machine learning. Typically, humans experts are trained on a system of dialogue act codes, then asked to use the codes to manually tag a set of transcripts. A dialogue act classifier is then developed and trained to automatically tag transcripts with some reasonably high degree of accuracy. As explained below, the availability of an effective dialogue act classifier also adds value to the real-time operation of a dialogue-based intelligent tutor, though it is insufficient on its own. Dialogue act classification is equally important to the retrospective analysis of massive databases of tutorial text, and perhaps equally inadequate, but for different reasons.

To explain why we say this, it is necessary to review the nature of tutorial dialogue itself. In Morrison & Rus (chapter 17), human learning is viewed as taking place in the context of an interaction between a learner and a "more knowledgeable other" (Vygotsky, 1978). The interaction typically involves a "joint attentional scene" (Bruner, 1983; Clark, 1996; Tomasello, 1999), in which at least one of the participants intends to acquire, or pass on, new knowledge or skill relevant to, or in some way represented by, the object(s) of joint attention. At the most granular level, the interaction consists of a series of back-and-forth *moves*, being actions that potentially affect the course of the interaction. Moves can include spoken or written utterances, gestures (e.g., pointing to an object), the act of drawing or writing, or physical actions, such as physical demonstration of tool use (e.g., see Keller & Keller,1996). In the case where a move has a particular intention, then it is interpretable as a *tactic*. Where moves are utterances, they may be identified as dialogue acts (Boyer et al., 2008; Rus et al., 2012; Stolcke et al., 2000), a category of *speech act* (Searle, 1969). Since most moves in pedagogical interactions can be assumed to have an intent, the terms *move, dialogue act,* and *tactic* are roughly synonymous, with one important distinction. Whereas a move, as a perceptible action, is apparent to both parties, the intent of the move and thus its tactical nature, remains largely hidden in the mind of the other. Further, to the extent that a tactical move represents a considered choice, then we can say that it must reflect an underlying strategy, which we define as an algorithm – sometimes called a "policy" (e.g., see Chi, VanLehn, Litman & Jordan, 2011) – for choosing from among possible moves (tactics), based on context, current circumstances, and an intended goal. Like tactics, strategies are hidden in the minds of interlocutors, and can therefore only be surmised. We also postulate the existence of something called a *metastrategy*, which is an algorithm for selecting from among available strategies. These different levels of analysis are illustrated in Figure 2.
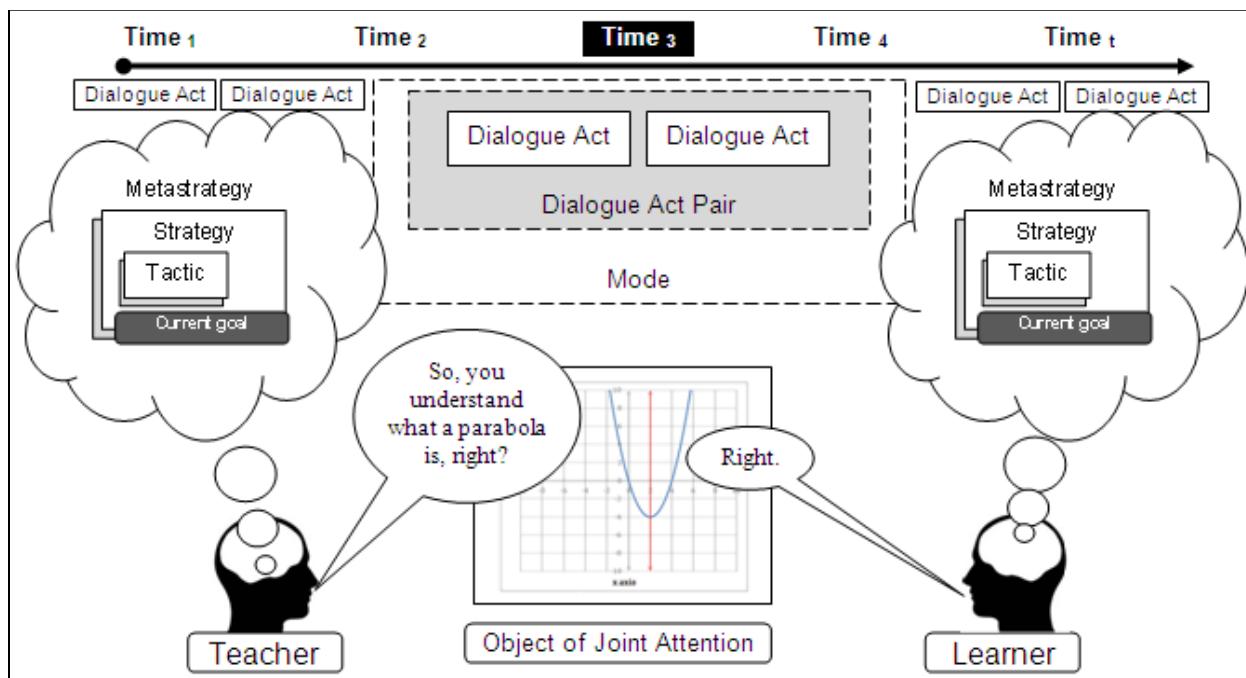
Figure 2. Anatomy of a tutorial dialogue.

Figure 3 illustrates these different levels of analysis over time during a tutoring session. Starting from the visible layers – where we can see objects of joint attention (or hear them referred to in the user's utterances) and hear/read utterances – we drill down to the invisible layers, which represent hidden mental constructs inside the heads of the interlocutors. Making sense of any given tutorial conversation involves developing a move-by-move theory about the intentions of both the learner and the tutor, which, in the

terms we are using here, means identifying utterances (moves) as particular dialogue acts, dialogue acts as tactics, the choice of tactics as reflecting strategies, and, where we suspect changes in strategy, the metastrategies that the interlocutors use to select from among strategies. A full accounting of a tutorial conversation, we argue, needs to take *all* of these levels of analysis into account. In other words, we can't say that we fully know what is going on in a tutorial conversation unless we have some sense of what is going on at all of these levels, and how, in particular, choices at one level affect choices at others. Understanding all of these levels is non-trivial and in some ways ill-defined, because grounding of meaning is inherently fuzzy. For example, the parties in the tutorial conversation often do not themselves agree about the goals of the current dialogue state. This leads to the illusion of discourse alignment, which is the misconception that the learner and tutor mutually understand why they are talking about a topic (Graesser, D'Mello & Person, 2009).
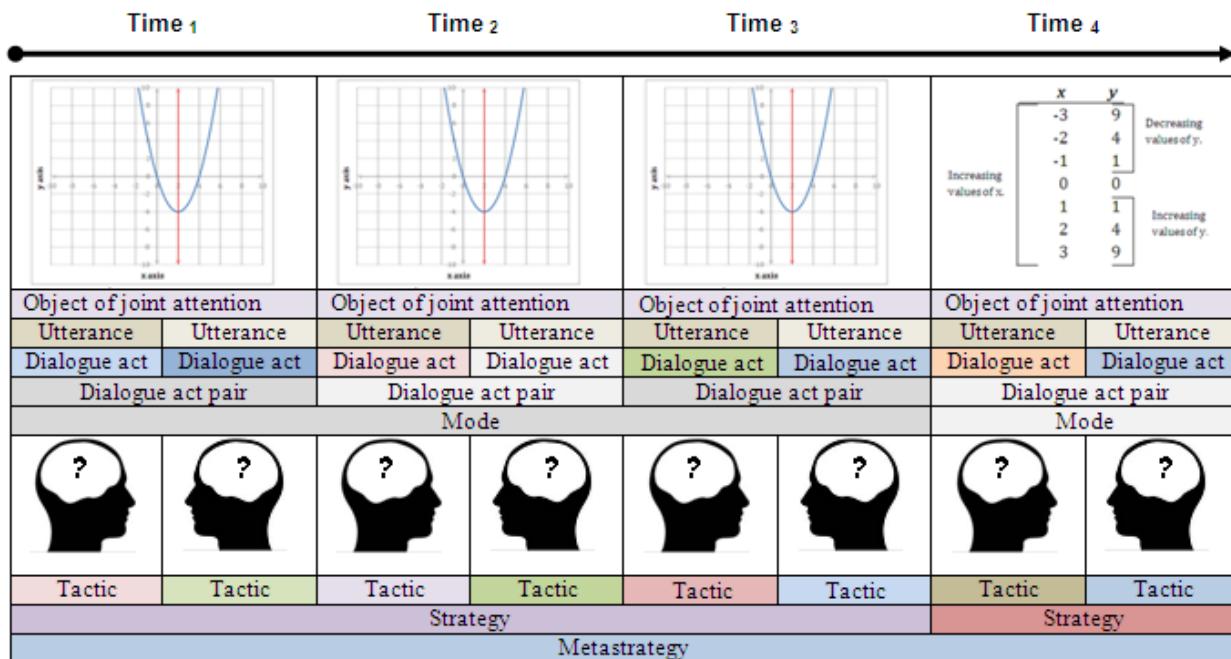
Figure 3. Levels of analysis[2].

Developing theories about the nature and function of dialogue act sequences is difficult, analogous in some ways to the problem of gene sequencing and the functional mapping of these sequences to the organism for which the DNA is a kind of blueprint. In this analogy, the dialogue acts are the nucleotides, and the dialogue act pairs are the amino acids. Clusters of dialogue acts and dialogue pairs – particular dialogue sequences – must, in some way, represent functional elements of the interaction, though how this works in particular cases will be far from immediately evident. The process of unlocking these mysteries, to the extent that researchers are able to do so, will likely involve a combination of top-down, bottom-up, and "sideways" processing, akin to solving a crossword puzzle by incrementally filling in the blanks, starting with the relatively easy bits, then using the resulting solution fragments as clues to help with the more difficult bits as a more complete solution gradually emerges. For example, having identified an utterance as a question, there is a probability that the utterance following it will be intended as an answer

---

[2] This model is problematic because it suggests that teachers and learner share the same underlying strategies and metastrategies, which is not the case. For example, while a tutor's strategy at a given point in the dialogue might be "assess the learner's current level of understanding," the learner's corresponding strategy might be "pretend to understand." The two also have different metastrategies.

to the question. Similarly, if a dialogue act classifier detects a series of what appear to be probes, prompts, and hints, these might be tentatively identified as tactics associated with a strategy aimed at eliciting "self-explanations." Conversely, if the working theory is that a particular dialogue segment represents an attempt to elicit self-explanations, then the dialogue classifier may be more likely to decide that a dialogue move that *looks* like a hint (has hint-like features) is in fact one.

Revisiting the LCC

Even careful readers who have stuck with us to this point may be confused about the nature of the systems we refer to. Are we talking chiefly about ITSs or human-human tutoring systems? And when we talk about the analysis of tutorial conversations, including the identification of dialogue acts, tactics, strategies, and metastrategies, do we refer to automatic machine learning algorithms, careful analysis of transcripts by human researchers, or some combination?

The answer to both questions is the same: Some combination. In fact, it seems likely that successful Internet-based learning systems of the future will be hybrid, multiagent systems that are capable of supporting and leveraging the work of useful combinations of both human and artificially intelligent teachers and researchers, in many different ways, and at many functional levels. An LCC agent is arguably a researcher in the very simple sense that it gathers data, analyzes it, then reports it in a form that others can use. For example, the LCC agent we describe at the outset of the chapter continuously monitors a human learner's dialogue moves in the course of a session with an intelligent tutor, characterizes these moves on the dimensions of *novelty* and *relevance* in real time, and then reports the results of its analysis in a language that other software agents can understand and may find useful. A dialogue act classifier agent, i.e., a module that analyzes a history of recent moves and, on this basis, reports its analysis in the form of a classification of the most recent move, is also a researcher in this sense. Further, as we have explained, the LCC agent will perform its job more effectively if it has access to output from its "colleague," the dialogue act classifier. In others words, what we are talking about here is a continuous stream of text data (the dialogue utterances, generated by some combination of humans and intelligent pedagogical agents), monitored by other agents, which then produce streams of messages of their own, which become available for use by still other agents. The result, illustrated in Figure 4, is the datastream we refer to in the title.
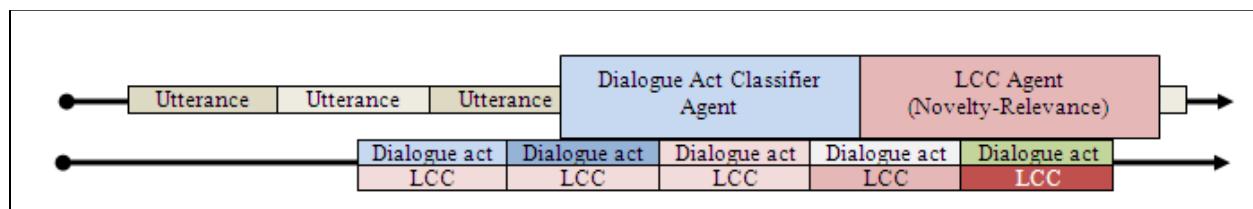
Figure 4. "Agents in the stream."

Conclusion

What, then, does this imply for micromodels such as an LCC? First, this view suggests that the LCC technique has value for evaluating a variety of tutoring transcripts, not just ones from systems such as AutoTutor. LCCs that track novelty and relevance are useful for assessing learner contributions intended to explain a concept or answer a question, even in human-human tutoring dialogues. The function of an LCC of this type is less straightforward when the ideal answer is not explicitly stored computationally, but it is still usable. Novelty, for example, is still a solid indicator of progress in a tutoring session:

students who repeat themselves are unlikely to be making progress. Relevance may also be reinterpreted in terms of the convergence between tutor and student dialogue contributions. Considering a tutorial dialogue as a Vygotskyian social process, a successful one should help the learner come to understand and be able to express concepts and relationships in the much the same way that the tutor does. An LCC can play a role at the level of dialogue act pairs and histories, considering the order of similar contributions (e.g., the learner states most of a concept first, which the tutor confirms, vs. learner simply repeats what the tutor says). This offers a metric for considering the roles and tactics in a wider variety of tutoring dialogs.

Second, the argument made here implies that LCC needs support from other tools to form a more robust learner micromodel for driving, or informing, a tutoring system. The role of the tutor's contributions also applies to an ITS. Failing to properly account for the knowledge imparted by the hints and prompts from the ITS likely gives undue credit to a learner in some cases. Conversely, as mentioned earlier, failing to classify dialogue acts before computing an LCC penalizes productive learning behaviors such as asking questions or expressing metacognitive statements about one's own learning state. Effectively, the simplest form of a novelty-relevance LCC does not model the intent (e.g., tactics or strategies) of the participants. For a system such as AutoTutor Lite, which dominates the selection of dialogue modes and topics, this issue can be sidestepped: assertions can be assumed because the tutor explicitly requests them. However, even for a related but more comprehensive system such as AutoTutor, a more complex learner micromodel is necessary.

This brings up a larger issue: The complexity of a learner micromodel depends significantly on the variety of tactics and strategies that the system accommodates. Studies of larger tutoring corpora should help map out a more comprehensive group of these tactics and strategies, for which effective micromodels can be developed. These micromodels should offer insights into the learner's knowledge state and also the overall effectiveness of a particular tactic or set of tactics. This is particularly important for advanced conversational tutors, which must use metastrategies to select among qualitatively different strategies and associated tactics (e.g., encouraging explanations vs. interactive lectures). Hybrid tutors and systems with multiple human tutors need effective micromodels for a different reason: knowing when to phone a friend. Much like doctors, tutors and tutoring systems should not specialize in all tutoring approaches and domains. For example, detecting wheel-spinning (Beck & Gong, 2013) with micromodels might be necessary to an effective referral system. In a tutoring-as-a-service paradigm, when a tutor concludes that an approach is not effective, the user could be referred to a personal learning assistant (Regan, Raybourn & Durlach, 2013), which might in turn refer the learner to a different learning resource, which will hopefully be more effective for that particular learner or topic.

Acknowledgments

References

Beck, J. E. & Gong, Y. (2013). Wheel-Spinning: Students Who Fail to Master a Skill. In H.C. Lane, K. Yacef, J. Mostow & P. Pavlik, (Eds.), *Proceedings of Artificial Intelligence in Education (AIED) 2013,* (pp. 431-440). Berlin, Germany: Springer.

Bruner, J. (1983). *Child talk.* New York: Norton

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P. & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145-182.

Chi, M., VanLehn, K., Litman, D. & Jordan, P. (2011). An evaluation of pedagogical tutorial tactics for a natural language tutoring system: A reinforcement learning approach. *International Journal of Artificial Intelligence in Education*, *21*(1), 83-113.

Clark, H.H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Douglass, S. A. & Mittal, S. (2013). A Framework for Modeling and Simulation of the Artificial. In *Ontology, Epistemology, and Teleology for Modeling and Simulation* (pp. 271-317). Springer Berlin Heidelberg.

D'Mello, S. K., Olney, A. & Person, N. (2010). Mining collaborative patterns in tutorial dialogues. *Journal of Educational Data Mining*, *2*(1), 1-37.

Forbes-Riley, K. & Litman, D. (2008). Analyzing Dependencies Between Student Certainness States and Tutor Responses in a Spoken Dialogue Corpus. In L. Dybkjaer & W. Minker (Eds.), *Text, Speech and Language Technology: Recent Trends in Discourse and Dialogue, Vol. 39*, (pp. 275-304). Dordrecht, Netherlands: Springer Netherlands.

Graesser, A. C., Chipman, P., Haynes, B. C. & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *Education, IEEE Transactions on*, *48*(4), 612-618.

Graesser, A. C., D'Mello, S. K. & Person, N. (2009). Meta-knowledge in tutoring. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice*. Mahwah, NJ: Erlbaum.

Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104 -137.

Graesser, A. C., Person, N. K. & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9,* 1-28.

Hu, X., Cai, Z., Han, L., Craig, S. D., Wang, T. & Graesser, A. C. (2009). AutoTutor Lite. In V. Dimitrova, R. Mizoguchi, B. Du Boulay & A.C. Graesser, A.C. (Eds.), *Proceedings of Artificial Intelligence in Education (AIED) 2009* (p. 802-802). Amsterdam: IOS Press.

Hu, X. & Martindale, T. (2008, January). Enhance learning with ITS style interactions between learner and content. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (Vol. 2008, No. -1).* National Training Systems Association.

Keller, C. M. & Keller, J. D. (1996). *Cognition and tool use: The blacksmith at work*. Cambridge University Press.

Landauer, T. K., Foltz, P. W. & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2-3), 259-284.

Miller, K. December, 2013. Personal communication with the first author.

Nye, B. D. & Morrison, D. M. (2013). Toward a Generalized Framework for Intelligent Teaching and Learning Systems: The Argument for a Lightweight Multiagent Architecture. In R. A. Sottilare & H. K. Holden (Eds.) *Artificial Intelligence in Education (AIED) 2013 Workshop on the Generalized Intelligent Framework for Tutoring (GIFT)*, (pp. 105-115). CEUR.

Regan, D., Raybourn, E. M. & Durlach, P. J. (2013). Learner Modeling Considerations for a Personalized Assistant for Learning (PAL). In R. Sottilare, A.C. Graesser X. Hu & H. Holden (Eds.), *Design recommendations for intelligent tutoring systems, Vol. 1 Learner Modeling,* (pp. 217-226). Orlando, FL: Army Research Laboratory.

Rosé, C., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A. & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International journal of computer-supported collaborative learning, 3*(3), 237-271.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). The Generalized Intelligent Framework for Tutoring (GIFT). Downloaded from www.gifttutoring.org on November 30, 2012.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

VanLehn, K., Jones, R. M. & Chi, M. T. (1992). A model of the self-explanation effect. *The journal of the learning sciences*, *2*(1), 1-59.

CHAPTER 20 – Intelligent Tutoring Support for Learners Interacting with Virtual Humans

Mark G. Core, H. Chad Lane, David Traum
University of Southern California

Introduction

Practicing face-to-face communication skills using virtual humans as partners is a challenging application area for ITSs. First, there is the inherent difficulty of the domain, but secondly there are challenges in communication between the virtual human and the ITS. Note, it is possible to use a human body to personify the ITS, e.g., Graesser, Jeon & Dufty (2008), but in this chapter, we use the term *virtual human* to refer solely to a virtual role-player that allows a learner to practice communication skills. The subject of this chapter is how to coordinate these two distinct activities: simulating a conversation partner (virtual human), and providing guidance and support to the learner (ITS). In particular, we examine several approaches to connecting the virtual human and tutor, including a specific example, RTXAI (Core, Lane et al., 2006; Core, Traum et al., 2006).

Communication skills are an "ill-defined" domain (Lynch, Ashley, Aleven & Pinkwart, 2006), given that the only directly observable aspects of problem solving are the utterances and nonverbal signals produced by the participants. Instructional design for these domains establishes a partial formalism, which can act as the basis for assessment and feedback. For example, Campbell et al. (2011) define strategies for young Naval officers counseling subordinates with personal or performance problems. These strategies are composed of actions and decision steps indicating the correct context in which to take those actions. An ITS can use these strategies to inform several components, including an expert model that represents expert competence, a learner model that tracks a learner's competence, and an instructional model that helps guide the learner toward expert status.

Figure 1 depicts a typical interaction between learner, virtual human, and ITS. The learner produces an utterance through spoken or typed input or through graphical user interface (GUI) manipulation (e.g., menu selection). The virtual human changes its internal state in response to this input, generates a response to the learner, and sends an update to the ITS containing the learner input, the virtual human response, and potentially details of its internal state changes. The ITS expert model then classifies the type of evidence (positive, negative, none) that the update provides for learner competencies on the modeled strategies. It is the responsibility of the ITS's learner model (see Figure 1) to maintain an estimate of the competency of the learner at any given time during learning (i.e., a measure of the learner's ability to enact the target strategies through the use of decision steps to take actions at appropriate times). Based upon the updated learner model, the instructional model provides appropriate support to the learner and may need to call upon the expert model to generate domain content (e.g., explanations). In addition to dynamic state updates, the ITS can also potentially reference static scenario data defining the initial state of the virtual human before the simulated conversation begins.
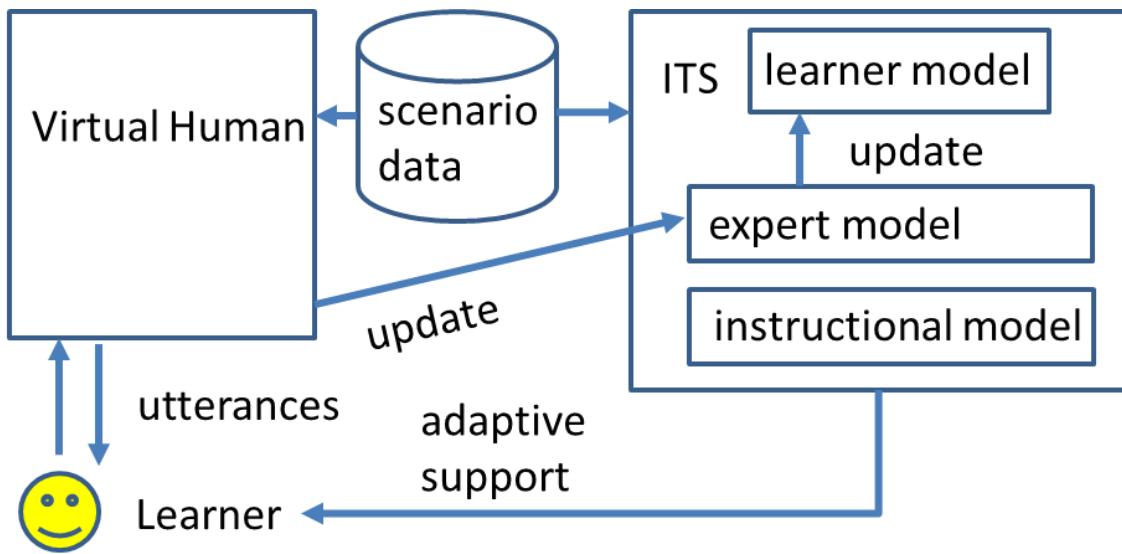
Figure 1. Learner model update in the context of virtual-human-based training.

It is tempting to extend the virtual human simulation to include the ITS capabilities. Such an architecture would minimize communication and coordination issues as well as providing the ITS full access to the simulation's state representation and reasoning capabilities. However, the resulting ITS would be tightly bound to the simulation, thus dramatically hindering reusability. To promote the use of ITS technology, efforts such as GIFT (Sottilare, Goldberg, Brawner & Holden, 2012) provide domain-independent tutoring software with the goal of allowing rapid development of ITSs in new domains and reuse of basic components. For example, such a toolkit means that developers can leverage existing code to update learner model probabilities or interface with hardware such as electroencephalogram (EEG) sensors.

If the ITS is not part of the virtual human simulation though, developers must provide the ITS with access to the necessary domain information to assess and support the learner. As we discuss in this chapter, a number of options exist to provide this information depending on the capabilities of the simulation. A particular problem is that some virtual human models are designed with only the end product of realistic behavior in mind, neglecting the importance of providing reliable forward-looking guidance and explanatory feedback (e.g., "The virtual human reacted negatively to you because…"). Developers of tutoring frameworks such as GIFT need to accommodate differing support for hints and explanations as well as aim to mitigate integration issues such as inconsistency between ITS hints and explanations, and virtual human behavior. To explore these issues, we start with a specific example, a virtual human serving as a practice partner for the skill of bilateral negotiation. We then describe other related work and end with discussion.

Example System: Reflective Tutor and Explainable Artificial Intelligence for SASO

Traum, Swartout, Marsella & Gratch (2005) describe a bilateral negotiation scenario called SASO implemented within a virtual human simulation. The learner plays the role of a military commander trying to convince a virtual doctor (see Figure 2) to move his medical clinic away from an insurgent-controlled neighborhood. To succeed, the learner must execute communicative strategies by speaking utterances to the character. The instructional design associated with the system includes many aspects of negotiation

such as convincing your counterpart to negotiate in the first place. For simplicity, we focus on the strategy of "building trust through solidarity" in this discussion.

Figure 2. Virtual doctor in SASO.

The virtual human is implemented in the Soar cognitive architecture (Laird, Newell & Rosenbloom, 1987) and includes models of dialogue and emotion encoded in production rules. In addition, speech recognition, speech synthesis, and non-verbal behavior generation (Lee & Marsella, 2006) allow communication between the learner and the virtual human. The virtual human's model of the world explicitly represents the preconditions and effects of physical actions. For example, treating patients requires medical supplies and moving an object has the effect of changing its location. Increased solidarity with the doctor will result from utterances such as offers that will help him achieve his goals (e.g., offering medical supplies helps the doctor achieve his goal of assisting patients). Decreased solidarity will result from utterances that threaten the achievement of the doctor's goals (e.g., offering troops to help move the clinic threatens the doctor's neutral status).

Core, Lane et al. (2006) and Core, Traum et al. (2006) describe a prototype ITS for SASO. This ITS interacts with the learner *after* the problem-solving activity using reflective tutoring tactics that have been shown to enhance learning and transfer (Katz, Allbritton & Connelly, 2003). We refer to a reflective tutoring session as an After Action Review (AAR). In addition to answering questions, the learner may also be asked to use a tool called explainable artificial intelligence (XAI), which is intended to reveal the reasoning and behaviors of the artificial intelligence (AI) system for the purposes of learning. The XAI tool allows a virtual human to answer questions about its beliefs and behaviors, in essence breaking character and being completely candid. Where possible and desirable, the virtual human should answer questions based upon the actual computations performed in its internal models. Virtual human beliefs may be different from the ground truth of the simulator, and the virtual human will generate answers using those beliefs.

The interface of the ITS and XAI system for SASO is shown in Figure 3. The upper-left quadrant shows a dialogue history; selecting an utterance (e.g., the highlighted "Yes" in the figure) means that the doctor will answer questions with respect to this time point in the scenario. The lower-right quadrant is an interface allowing learners to select questions to ask, and the upper-right quadrant is a history of the questions asked, and the answers given. As shown in the bottom-left quadrant, the ITS reviews the session with the learner through an interactive dialogue. The ITS gives learners an "investigation goal,"

tracks their progress, and gives hints as necessary (e.g., "We first need to know why the negotiation failed."). The ITS is called RTXAI, Reflective Tutor (i.e., a tutor that supports reflection) and XAI.
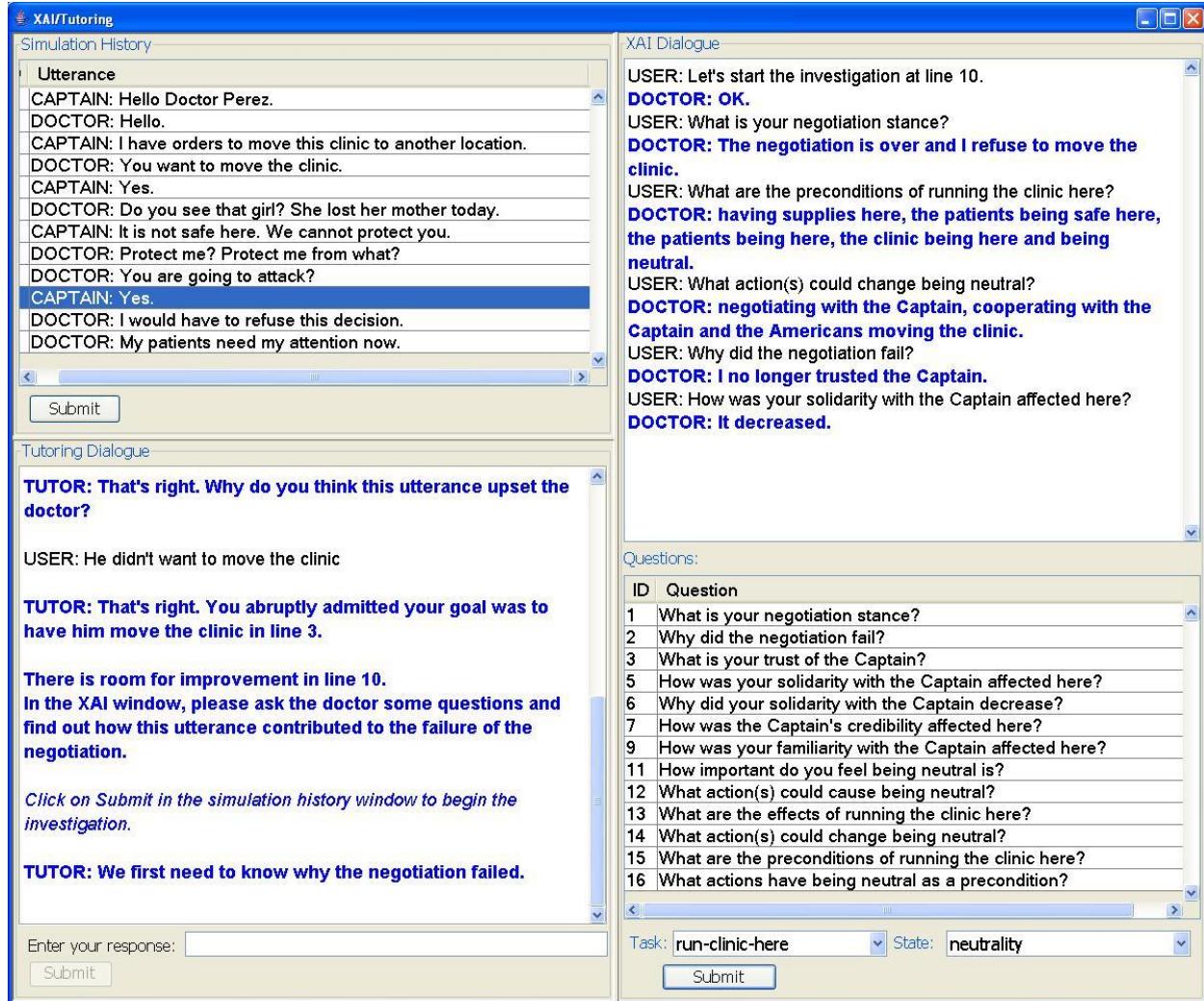
Figure 3. ITS Interface for RTXAI for SASO.

Design Considerations for an ITS/Virtual Human Interface

RTXAI is a separate system implemented outside of the virtual human simulation, and Figure 1 roughly captures the relationship between RTXAI and the virtual human simulation. As an example, we give the details of the process that occurs when the learner utters "Yes" confirming that the Army is going to attack the local insurgents (see Figure 3). The virtual human updates RTXAI with the information that solidarity has decreased and indicates the specific production rule triggering this change. No explicit learner model is used here, but if a learner model were introduced then expert model rules would trigger a decrease in the estimated competency of the learner in the "building trust through solidarity" strategy.

In Figure 3, the learner is using XAI to question the doctor about what happens after the learner confirms the Army is going to attack. The tutor has guided the learner to uncovering that solidarity has decreased by hinting, "We first need to know why the negotiation failed." Once a variable change is revealed, an

associated "why" question is unlocked and appears in the menu. The learner investigates the loss of trust by asking about solidarity. If the learner continued by asking, "Why did your solidarity with the Captain decrease?" then XAI would translate the associated production rule into an English explanation (e.g., "The Captain is committing to performing an undesired act").

To build RTXAI for SASO, we considered four possible approaches.

Approach 1: Some simulations support post-hoc explanation (Johnson 1994; McAlinden, Gordon, Lane & Pynadath, 2009), in which case it is the responsibility of the simulation to produce causal information for dynamic state updates. ITS authors would then only need to build natural language generation resources to convey the explanations to the learner. This is the best possible situation as it avoids the need to import scenario data and minimizes the need to duplicate the reasoning capabilities of the virtual human simulation.

Approach 2: If the simulation does not support explanation, the next best option is for the ITS to import the scenario data and calculate causality itself. For example, consider a virtual human selecting a plan to achieve a goal. The ITS can reason about what elements of the initial conditions were necessary for the plan and under what conditions the virtual human might have chosen a different option. As long as the ITS's reasoning process follows the same methodology as that of the virtual human, there should be no discrepancy between what happened in the simulation and the ITS's explanation. A drawback of approach 2 is that it is likely to require redundancy between the ITS and the supported simulation (i.e., the ITS performing identical tasks that occur in the simulation).

Approach 3: If the ITS cannot use the scenario data directly, then it may be able to use scenario data annotated with information about causality. For example, production rules encoding models of behavior can be difficult to interpret since the left-hand side of the rules may include operations such as binding variables in addition to the triggering conditions for a behavior. Similarly, the right-hand side of these production rules can include operations such as internal bookkeeping in addition to the actual effects of the action. An author could annotate important variables on the left- and right-hand sides of production rules such that chains of causality could be built post-hoc to explain virtual human behavior.

Approach 4: The last option is to hand build a separate representation of virtual human behavior for the ITS. Sometimes this ITS model of the virtual human is specific to the scenario being modeled since this is simpler than building a general causal model of the virtual human's behavior. The benefit of the approach is that it makes no assumptions about the virtual human implementation. The drawbacks of the approach are that any future changes in the virtual human or scenario data must be reflected in the ITS model, and that causal information missing in the virtual human data must be encoded in consultation with scenario authors and subject matter experts. Because the behavior model of the virtual human is not directly driving ITS guidance and XAI explanations, there is no guarantee that such guidance and explanations will be completely consistent with virtual human actions.

The virtual human in the SASO scenario (Traum et al., 2005) does not support post-hoc explanation so we were limited to options 2-4 in implementing RTXAI for SASO. As noted above, it is difficult to reason causally using production rules, because these rules mix low-level implementation details with high-level concepts (e.g., triggering a behavior, changing a belief). However, physical actions were represented in SASO in such a way that RTXAI could reason about them directly (i.e., approach 2), although as noted in Core, Lane et al. (2006), some human intervention was necessary in the import process. For mental actions and concepts (e.g., trust), we built a separate representation of the relevant production rules including their relationship to the instructional design (e.g., solidarity) and English translations of the rules for XAI. Potentially we could have used approach 3 for this aspect of the SASO

virtual human, but time limitations led us to use approach 4, which enabled quick development of a prototype.

To implement XAI capabilities, we set up a relational database to store and retrieve simulation data, and defined a series of question templates. Figure 3 shows the XAI input GUI in the bottom-right quadrant. In this GUI, the learner selects fillers for the two template variables, task and state, changing the questions presented in the menu above. The questions are represented internally using a logical form, and questions and answers are translated into English using domain-specific natural language generation resources. The SQL query used to retrieve the answer to the question is generated automatically from the question representation.

Related Research

Virtual-human-based training systems for conversational skills range from fielded systems, e.g., Johnson (2010), to lab-based prototypes such as the RTXAI system (Core, Lane et al., 2006; Core, Traum et al., 2006) described above. Getting systems into the hands of learners means that reliability and authorability become critically important. The system must behave predictably if developers are to test all of the possible learner experiences. Support must be provided to authors such that a sufficient quantity of scenario data can be created to provide learners with the necessary practice opportunities. These constraints discourage developers from implementing virtual humans that reason about their responses using models of tasks, dialogue, and emotion because it can be difficult to predict their behavior. Also, authors must consider how their changes to these models not only expand the possible dialogues but also modify how the virtual human responds to previously seen learner utterances.

Thus, fielded systems tend to use simple and predictable models of dialogue. In the area of virtual patients, Rossen & Lok (2012) characterize current systems as simply matching learner inputs to pre-authored question/answer pairs. Knowledge-lean approaches are also used in training face-to-face communication skills in the military in systems such as INOTS (Campbell et al., 2011), BiLAT (Kim et al., 2009), and VECTOR (Barba et al., 2006). In these systems, learners select conversational actions from menus navigating a tree-based conversation space. In BiLAT and VECTOR, variables store the effect of learner choices so that they can influence later parts of the conversation (e.g., a negotiation may never succeed until the trust variable exceeds a threshold). Because the virtual human is simply following a tree and making decisions based on a handful of variables, authors must manually annotate each choice with its relationship to the strategies being practiced (e.g., is this choice evidence that the learner has mastered the strategy?). This annotation amounts to building a separate representation for the ITS since the tree-based model contains no causal information to explain the effect of learner actions.

In cultural training, such hard-coded approaches make it difficult to reuse data when expanding a system to train learners to deal with multiple cultural environments. Sagae, Ho & Hobbs (2012) describe an effort to modularize the resources of training systems built by the company Alelo. To avoid hard coding culture into models, Sagae et al. develop reusable principles such as greetings can be used to establish friendship. For a particular language and culture, words such as "Salaam Alaikum" can be labeled as a greeting. In addition to Sagae et al. (2012), there have been other efforts to model culture modularly (Buede, DeBlois, Maxwell & McCarter, 2013; Silverman et al., 2012; Solomon, van Lent, Core, Carpenter & Rosenberg, 2008; Zielke & Linehan, 2009). The work of Silverman et al. (2012) is especially notable because their characters have the ability to discuss their internal models with the player meaning that an ITS could perhaps interface with the simulation to extract this information to provide assessment and explainable AI capabilities.

This trend toward reusable models of culture and dialogue is likely to continue. Although hard coding culture and dialogue into scripts does not require technical expertise and results in predictable behavior, such scripts are not reusable and require each piece of dialogue to be linked to the instructional design. In general, the authoring burden is high. Many lines of realistic dialogue must be written, and authors must keep track of the many paths through the conversation maintaining consistency as well as ensuring sufficient opportunities for learners to practice the target skills.

Discussion

There is increasing recognition of the importance of interpersonal skills training and the use of virtual humans (more generally, "virtual training environments") in training (Department of the Army, 2011). An ITS is an important part of such practice environments, and a tutoring framework such as GIFT (Sottilare et al., 2012) could aid ITS developers building interfaces to virtual human simulations. Building such interfaces is complex because the expert model of the ITS needs not only to label learner actions as correct or incorrect, but also identify causal links between learner actions and the behavior of the virtual human. In this chapter, we have discussed different approaches to interfacing an ITS with a virtual human:

- **Virtual human provides causal information about simulated conversation**. Here interfacing chiefly concerns communication with the virtual human and translating information received into natural language that can be understood by the learner.

- **ITS imports static scenario data and performs its own causal reasoning**. In this case, authors may need to annotate scenario data such as production rules with information about causality (e.g., which parts of the rules correspond to the triggers and effects of a behavior). An additional task in this approach is translating the simulation-specific scenario data to the ITS's internal knowledge representation.

- **ITS maintains its own model of the virtual human and its behavior**.

A tutoring framework can provide useful tools for developers of an ITS for a virtual human, in particular, knowledge representation and reasoning components. Depending on the type of interface used, the ITS can use these representation and reasoning capabilities to import static scenario data and dynamic trace information, as well as generate explanations and predictions about virtual human behaviors. These explanations and predictions can then be used to provide guidance to learners.

As discussed in the Related Work section, developers of virtual humans use a wide variety of models and it is a challenging task to provide simulation-independent interface support. However, it is crucial that an ITS for a virtual human be able to help learners who may have difficulty recognizing the effects of their actions in simulated conversations. Communication skills training is necessarily abstract with respect to a specific situation and the actual words to be spoken. In SASO (Traum et al., 2005), one of the general strategies is to make offers that help the virtual doctor achieve his goals and avoid as much as possible requests that threaten the achievement of the doctor's goals. However, the learner may not know the second- and third-order effects of their actions and how they might relate to the doctor's goals. The learner may also think they performed a particular conversational action (e.g., offering help moving the clinic), but the virtual human interpreted the learner in a different way (e.g., a threat to the doctor's neutrality). Also problematic are "curve balls," where a correct action results in a negative response from the virtual human, or an incorrect action triggers a surprisingly neutral or positive result. To avoid development of misconceptions, it is important that learners recognize the weak or non-existent link between their actions and the curve-ball responses.

In fact, as discussed by Lane & Wray (2012) and Wray et al. (2009), the ideal interface between virtual human and ITS should allow the ITS to decide whether the virtual human gives a curve-ball response because it should be timed appropriately based on the learner's performance and delivered when they are ready. Generally, developers of virtual humans should consider providing interfaces that allow external ITS software to perform experience manipulation. Experience manipulation could be used to adapt the virtual human's behavior to the individual learner who may need encouragement (e.g., a forgiving virtual human who is upbeat and positive), feedback (e.g., character may explain why it thought the learner was being offensive), challenge (e.g., a difficult character who generally has negative responses), or variety (e.g., force the learner to exercise different communicative strategies).

References

Barba, C., Deaton, J. E., Santarelli, T., Knerr, B., Singer, M. & Belanich, J. (2006). Virtual environment composable training for operational readiness (VECTOR). In Proceedings of the Army Science Conference.

Buede, D., DeBlois, B., Maxwell, D. & McCarter, B. (2013). Filling the need for intelligent, adaptive non-player characters. In Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).

Campbell, J., Core, M., Artstein, R., Armstrong, L., Hartholt, A., Wilson, C., . . . Yates, K.A. (2011). Developing INOTS to support interpersonal skills practice. In Proceedings of the IEEE Aerospace Conference.

Core, M., Lane, H. C., van Lent, M., Gomboc, D., Solomon, S. & Rosenberg, M. (2006). Building explainable artificial intelligence systems. In Proceedings of the Eighteenth Conference on Innovative Applications of Artificial Intelligence (IAAI).

Core, M., Traum, D., Lane, H. C., Swartout, W., Gratch, J., van Lent, M. & Marsella, S. (2006). Teaching negotiation skills through practice and reflection with virtual humans. SIMULATION, 82 (11), 685-701.

Department of the Army. (2011). The U.S. Army Learning Concept for 2015 (TRADOC Pam No. 525-8-2). Training and Doctrine Command, Fort Monroe, Virginia.

Graesser, A. C., Jeon, M. & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. Discourse Processes, 45, 298-322.

Johnson, W. L. (1994). Agents that learn to explain themselves. In Proceedings of the Twelfth National Conference on Artificial Intelligence.

Johnson, W. L. (2010). Serious use of a serious game for language learning. International Journal of Artificial Intelligence in Education, 20 , 175-195.

Johnson, W. L., Sagae, A. & Friedland, L. (2012). VRP 2.0: Cross-cultural training with a hybrid modeling architecture. In D. D. Schmorrow & D. M. Nicholson (Eds.), Advances in Design for Cross-Cultural Activities (p. 307-316). Boca Raton: CRC Press.

Katz, S., Allbritton, D. & Connelly, J. (2003). Going beyond the problem given: How human tutors use post-solution discussions to support transfer. International Journal of Artificial Intelligence in Education, 13, 79-116.

Kim, J., Hill, R.W., Durlach, P., Lane, H. C., Forbell, E., Core, M., … Hart, J. (2009). BiLAT: A game-based environment for practicing negotiation in a cultural context. International Journal of Artificial Intelligence in Education, Special Issue on Ill-Defined Domains, 19(3).

Laird, J.E., Newell, A. & Rosenbloom, P.S. (1987). Soar: An architecture for general intelligence. Artificial Intelligence, 33(1), 1-64.

Lane, H. C. & Wray, R. (2012). Individualized social and cultural learning with virtual humans. In P. Durlach & A. Lesgold (Eds.), Adaptive technologies for training and education (p. 204-221). New York: Cambridge University Press.

Lee, J. & Marsella, S. (2006). Nonverbal behavior generator for embodied conversational agents. In Proceedings of the International Conference on Intelligent Virtual Agents.

Lynch, C. F., Ashley, K. D., Aleven, V. & Pinkwart, N. (2006). Defining "ill-defined domains"; a literature survey. In Proceedings of the ITS 2006 Workshop on Intelligent Tutoring Systems for Ill-Defined Domains.

McAlinden, R., Gordon, A., Lane, H. C. & Pynadath, D. (2009). UrbanSim: A game-based simulation for counterinsurgency and stability-focused operations. In Proceedings of the AIED 2009 Workshop on Intelligent Educational Games.

Rossen, B. & Lok, B. (2012). A crowdsourcing method to develop virtual human conversational agents. International Journal of Human-Computer Studies, 70, 301-319.

Sagae, A., Ho, E. & Hobbs, J. R. (2012). Efficient cross-cultural models for communicative agents. In Proceedings of the Second International Conference on Cross-Cultural Decision Making.

Silverman, B. G., Pietrocola, D., Nye, B., Weyer, N., Osin, O., Johnson, D. & Weaver, R. (2012). Rich socio-cognitive agents for immersive training environments: case of nonkin village. Autonomous Agents and Multi-Agent Systems, 24(2), 312-343.

Solomon, S., van Lent, M., Core, M., Carpenter, P. & Rosenberg, M. (2008). A language for modeling cultural norms, biases and stereotypes for human behavior models. In Proceedings of the Seventeenth Conference on Behavior Representation in Modeling and Simulation (BRIMS).

Sottilare, R. A., Goldberg, B. S., Brawner, K. W. & Holden, H. K. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).

Traum, D., Swartout, W., Marsella, S. & Gratch, J. (2005). Fight, flight or negotiate: Believable strategies for conversing under crisis. In Proceedings of the Fifth International Working Conference on Intelligent Virtual Agents.

Wray, R., Lane, H. C., Stensrud, B., Core, M., Hamel, L. & Forbell, E. (2009). Pedagogical experience manipulation for cultural learning. In Proceedings of the AIED 2009 Workshop on Culturally Aware Tutoring Systems.

Zielke, M. A. & Linehan, T. E. (2009). The first person cultural trainer. In Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).

SECTION IV

Instruction and Scaffolding

A. Graesser, Ed.

CHAPTER 21 – **Guided Instruction and Scaffolding**

Arthur Graesser
University of Memphis

Introduction

Instructional strategies to improve learning and motivation have been identified in many different fields: education, educational psychology, cognitive science, learning sciences, computer-based training, ITSs – the list goes on. Instructors, tutors, and mentors typically have intuitions on what learning strategies work for particular types of learners. Pedagogical theorists advocate particular strategies and offer elaborate explanations on why they should be effective. Empirical scientists collect data from different categories of learners in different instructional conditions in order to gather informative data on whether strategies truly facilitate learning and motivation. Computational scientists build computer systems that precisely implement strategies and track data moment-by-moment in rich detail. Assessment experts develop psychometrically valid tests of learning and motivation in order to evaluate the value of various strategies with solid measures. Policy makers decide on education programs and curricula that emphasize particular strategy profiles, based on the reports of the above stakeholders and, of course, the available budgets.

Everyone has an opinion on the value of various strategies. It is likely that all of these opinions are both correct and incorrect. That is, they are correct for some learners in some contexts with some strategy delivery systems. But they are not correct for many others. From the standpoint of science and good engineering, our goal is to identify and tune the precise conditions, content, and context when a strategy works. Our approach to evaluating a strategy lies in precise theoretical specification, computational implementation, and empirical validation. Good ideas and vague intuitions come a dime a dozen. Analytical precision with evidence is the gold standard for GIFT.

The chapters in this section focus on two foundational constructs associated with strategies, namely scaffolding and the ZPD. The contributors to the five chapters agree that these constructs need a more precise specification that researchers and practitioners need to firmly acknowledge. These constructs should not be confused with mere support by humans or technology. So what is scaffolding and ZPD? At the risk of oversimplifying the constructs, a few guiding ideas should be helpful, but the discussions evolve in the five chapters. A scaffold can be a book, diagram, tutor, computer technology, or any other material or processing facility that can allow some degree of control by the learner. A scaffold is not a rigid activity that unfolds mechanically without adapting to or allowing some control by the learner. A scaffold is not a wide open world that allows learners to do whatever they want. Everyone agrees with these claims, but the big question is what counts in between. That is what requires precise specification.

ZPD algorithms start out identifying what learning can be accomplished without the scaffold (learning from L_o) versus with the scaffold (learning from L_s). The scaffold is within the learner's ZPD if (L_s - L_o). There is also the claim that the materials should not be too easy or too difficult, but just right, in the ZPD. These claims most agree with, but the challenge is how to do this and be more precise in specification. The chapters in this section identify more specific scaffolding content and processes that give more informative guidance to developers of advanced learning environments. Some authors advocate a more detailed theoretical analysis of the scaffolding process. We need to say more than expressing the Goldilocks principle (i.e., the materials should not be too easy or too difficult, but just right) and that learners need to go through an evolution of modeling, scaffolding and fading. There needs to be more concrete and quantitative guidance on how to accomplish this. The guidance needs to accommodate specific affordances of the intervention materials (text, diagrams, computer, people) and the multidimensional learner profile.

Chapters

Chapter 22 by Holden and Sinatra reiterates the importance of defining scaffolding at a sufficiently precise level to be helpful to GIFT. The focus of this chapter is on self-regulated learning strategies and metacognition, two critical learner characteristics that will be needed in the future. Many believe that future learners are expected to take more responsibility for their learning and mastery of important skills. Computer environments, as well as human tutors, vary in their support for helping students develop self-regulation and metacognitive strategies. Systems with conversational agents can provide some guidance but hybrid systems between human tutors and ITS are anticipated in the future.

Chapter 23 by Schnotz and Cade discusses strategies of presenting and coordinating information sources in multimedia. Multimedia has different forms of representation (text, diagrams) and different modalities (visual, auditory) that somehow need to be selected and combined in learning technologies. Which media and modalities help the learner most? It depends on the background knowledge of the learner and the complexity of the multimedia. What conditions would create a split attention between modalities? It depends on the learner in addition to some general constraints of human cognition. The chapter presents production rules and a network architecture that generates predictions on what ensemble of multimedia features should help learning for particular classes of learners.

Chapter 24 by Durlach identifies different classes of advanced learning technologies with respect to the levels of adaptivity and strategic intervention. Aside from computer-based training systems that are not adaptive to the learner, there are those that assign the next problem in a way that is sensitive to the learner's achievement and abilities (macro-adaptation), those that generate moves that adapt to actions and decisions of learners within a problem (micro-adaptivity), those that do both, and those that combine global traits of the learner with both macro and micro levels. Durlach aptly points out that evidence is scarce for the benefits of the more sophisticated forms of adaptivity and associated strategies. Her observations should provide a blueprint for future ITS research and development.

Chapter 25 by Rus, Conley, and Graesser pushes the limits of the depth and grain-size of adaptive ITS strategies. They defined a hierarchical dendrogram that spans many levels of instruction granularity in learner modeling, strategies, and tactics. Their development of DeepTutor incorporates learning progressions that handle learner modeling, macro-adaptation that selects next problems to work on, and micro-adaptation that generates feedback, hints, and prompts with natural language interaction. This achievement requires computational breakthroughs in computational linguistics, artificial intelligence, discourse, and cognitive science.

Chapter 26 by Olney begins with a historical analysis of the constructs of strategy and ZPD. He aptly points that these constructs have been used very informally and loosely over the decades. It is also proposed that there is value in precise theories of scaffolding that can identify specific steps and processes at sufficient detail that they could be applied to the GIFT architecture. The chapter also discusses the importance of making abstract ideas and processes visible. Some of the original theories of scaffolding handled perceptual-motor procedures but the various skills of today (like algebra) are abstract. Nevertheless, the scaffolding steps and processes can still be applied with the aid of computer technologies.

Implications for GIFT

The wisdom in these chapters provides both general and specific recommendations on the strategies incorporated in the advanced learning environments developed with GIFT. Both ZPD and scaffolding require adequate learner modeling so that the application of any strategy is aligned with the relevant parameters of the learner profile (e.g., skill level, prior knowledge, mastery of particular topics, motiva-

tion, affect). The strategy needs to be represented in sufficient detail that it can be implemented in a computer system and can accommodate the specific conditions for successful application. Future ITSs need to incorporate more sophisticated components with macro-adaptivity and micro-adaptivity that are part of the holy grail of the ITS enterprise (but often not implemented in existing systems). There needs to be more discriminating empirical tests of the added value of particular intelligent components on learning and motivation.

At some point in GIFT research and development, there needs to be greater attention to the grain size of learner modeling metrics, the precision of the strategy parameters, and the dosage of the strategy interventions. At early stages in the research and development (R&D) cycle, it makes sense to have crude metrics and precision (i.e., low, medium, versus high) and large dosages. At later stages, there will invariably be enhancements in granularity, precision, and appropriate dosage. Budgets and practical concerns will demand such quantification.

CHAPTER 22 – A Guide to Scaffolding and Guided Instructional Strategies for ITSs

Heather K. Holden and Anne M. Sinatra
U.S. Army Research Laboratory (ARL), Human Research and Engineering Directorate (HRED),
Simulation Training and Technology Center (STTC)

Introduction

Providing the optimal level of desired challenge and flow during instruction can be difficult for any type of educational/learning setting (i.e., traditional classroom-based instruction, one-to-one peer tutoring, online/hybrid learning environments, etc.). The delicate art of scaffolding has been a key area of interest among educational researchers and psychologists over the last 30 years due to its advantage in promoting learning educational environments. According to Wood, Bruner, and Ross (1976), scaffolding was originally defined as an "adult controlling those elements of the task that are essentially beyond the learner's capacity, thus permitting him to concentrate upon and complete only those elements that are within his range of competence" (p. 9). With the scaffolding process, learners are able to carry out a task, achieve a goal, or solve a problem that they could not on their own accord. Although not originally connected, scaffolding is often explained in terms of Vygotsky's ZPD, which suggests that learning is a social process that elevates the learner from their actual development level to a "higher level of potential development as determined through problem solving under adult guidance and in collaboration with more capable peers" (Vygotsky, 1978, p.86). A scaffold is considered to be a temporary support and is gradually removed ("faded") once the learner reaches their potential and becomes sufficient at performing the task on his/her own. The overall notion is that a novice can best reach his/her learning potential with scaffolding provided by a more knowledgeable adult/expert, i.e., a parent, teacher, tutor, peer (Lajoie, 2005; Puntambekar & Hubscher, 2005). Consequently, the role of the adult expert is the most important role in scaffolding, and the instructional relationship that the adult expert has on a learner's development is crucial to achieving deep learning.

Scaffolding is considered to be an interpersonal process in which both the teacher and the learner actively participate to build common understanding through their communication exchanges; consequently, the learner learns from the perspective of the teacher (Stone, 1993). One of the primary issues with scaffolding is the lack of clarity in its definition, conceptualization, and standardization of research. Some researchers suggest that application of scaffolding has been too broadly used in educational and psychological research. As a result, current research has deviated from the original scaffolding theory. Pea (2004) stated that "the concept of scaffolding has become so broad in its meaning in the field of educational research and learning sciences that it has become unclear in its significance" (p. 423). Puntambekar and Hubscher (2005) concurred, suggesting that "the scaffolding construct is increasingly being used synonymously with support" (pg. 1). Clearly, the ambiguity of the scaffolding metaphor has become a dilemma.

The purpose of this literature review is to provide an overview of the research on scaffolding in computer-based learning environments and identify its applicability to the future development of ITSs. In this literature review, we present 1) the original conception of scaffolding and how it pertains to traditional implementations of classroom and one-to-one human tutoring (the historical perspective); 2) the evolved conception of scaffolding as it pertains to computer-based learning environments (the current perspective); and 3) recommendations for future scaffolding research and ITSs.

The Original (Historical) Conception of Scaffolding

In the original conception of scaffolding, the role of a single knowledgeable adult expert, notably a parent or teacher, is the one of the most critical components. According to Wood, Burner, and Ross (1976), the adult expert could provide six types of support to the student: recruiting the child's interest, reducing the degrees of freedom by simplifying the task, maintaining direction, highlighting the critical task features, controlling frustration, and demonstrating ideal solution paths. The adult is a representation of the following roles: a domain expert; a knowledgeable facilitator of the skills, strategies, and processes required for effective learning; a motivator for the learner to support him/her to achieve the desired goal; and a provider of modeling, hints, and questions for the learner to reflect on the presented information (Puntambekar & Hubscher, 2005; Wood, Bruner & Ross, 1976). Stone (1998) considers the adult's role as a combination of perceptual, cognitive, and affective elements (Stone, 1998).

Intersubjectivity, or a shared understanding of the task to be performed, is a crucial element of the scaffolding process. However, van de Pol, Volman & Beishuizen (2010) conducted a literature review of research on scaffolding in the classroom and found that, despite the different definitions of scaffolding, there are three common key characterizations of the scaffolding process: 1) contingency (ongoing diagnosis and calibrated support), 2) fading, and 3) transfer of responsibility (van de Pol, Volman & Beishuizen, 2010). Van de Pol, et. al. (2010) defines *contingency* as the calibrated support and responsiveness by the adult expert (i.e., teacher). The teacher cautiously responds/adapts contingently based on an understanding of the task being performed, the student's performance, and diagnostic strategies to identify the student's current level of learning. The importance of this *ongoing diagnosis* of the student's current level of understanding has been considered vital to scaffolding by many researchers. Contingency (ongoing diagnosis) has also been referred to as dynamic assessments, which is defined as the "moment-by-moment assessment of learners while they are in the process of problem solving for the purpose of making informed decisions about feedback" (Lajoie, 2005, p. 545). *Fading* is considered the gradual reduction of the scaffold, depending on the learner's competence and level of development. The fading process is directly related to the third characterization, *transfer of responsibility*. As contingent fading is implemented, the responsibility for learning is transferred when a student takes increasing learner control of the performance of the task (van de Pol et al., 2010). Although the definition of scaffolding is often confused with support, these three characterizations must be implemented for true scaffolding.

Conceptualization of Scaffolding

There is no formal or universally accepted framework for analyzing and classifying scaffolding strategies, but a few researchers have tried to develop such a framework. Based on the literature, two core questions need to be addressed in order to produce a scaffolding strategy: *What to scaffold* and *how to scaffold* (Azevedo & Jacobson, 2008; Lajoie, 2005; Pea, 2004; van de Pol et al., 2010). Determining what to scaffold can be difficult and requires researchers to be cautious when designing studies evaluating the effectiveness of learning and scaffolding strategies. The sub-questions to answer include the following:

1. Is the focus on scaffolding the topic/domain or the learning processes underlying domain learning (e.g., metacognitive processes, problem solving, and self-regulatory processes) or both content and processes? (Azevedo & Jacobson, 2008)

2. What are the goals or intentions (cognitive, metacognitive, or affective) of scaffolding? (van de Pol et al., 2010)

3. Why scaffolding?

Ascertaining and responding to knowledge of the learner's individual differences, such as level of prior knowledge, developmental level, and domain expertise, is an important element in understanding what to scaffold. For example, learners with low prior domain knowledge and domain expertise may require both scaffolding of the domain content and process, whereas students with higher prior knowledge may only need process scaffolding (Azevedo & Jacobson, 2008).

Determining how to scaffold is the second core question in constructing a scaffolding strategy and pertains to how the scaffolding is going to take place. The sub-questions to answer include the following:

1. What are the tools or methods (feeding back, hints, instructing, explaining, modeling, and questioning) of scaffolding? (van de Pol et al., 2010)

2. When is a scaffold administered or made available to the learner during learning and when is the scaffold faded? (Azevedo & Jacobson, 2008).

Van de Pol (2010) suggests that any combination of scaffolding means (how to scaffold) and scaffolding intentions (what to scaffold) can be constructed as a *scaffolding strategy*. This conceptualization of scaffolding tends to be the most common in the early scaffolding literature. Rosenshine and Meister (1992) suggested that a scaffold may be either a *tool*, where a scaffolding device, such as a cue card, is provided to the learner, or a *technique*, such as a strategy that the teacher implements in order to support a learner (Rosenshine & Meister, 1992; Yelland & Masters, 2007). Moreover, Yelland & Masters (2007) identified the conception of scaffolding as a temporal component, "with respect to both *type* and *extent* of scaffolding provided" (p. 364).

A primary issue with the original conception of what to scaffold is the predominate focus of scaffolding pertaining to learners' cognition and performance, but with a lack of focus on learners' motivation and affect. Van de Pol et al. (2010) discovered that the scaffolding of cognitive and metacognitive activities is the most researched compared to the scaffolding of learner's affect. They also found that the means (e.g., how to scaffold) of modeling and questioning are the most researched, predominately with a focus on learners' cognitive activities. However, research has shown that cognitive, conative (i.e., beliefs, motivation, etc.), and affective structures have a strong relationship to learning (D'Mello, Lehman & Graesser, 2011; Lajoie, 2005). Yelland and Masters (2007) reported that learners need affective scaffolding of varying amounts to not only keep them on task, but also to encourage them to achieve higher levels of thinking when engaged with a variety of learning activities. Pea (2004) stated that scaffolding must consider both cognitive as well as motivational aspects of learning. Such factors become more important for scaffolding in computer-based learning environments. The current disconnect between the theoretical and implementation of scaffolding is primarily due to the change in dynamics of the teacher-learner interaction and the learning environments. The adult expert in the scaffolding process can now be a human (teacher, peer, or parent) or a nonhuman (virtual tutor or pedagogical agent). The learning environments can now be one-to-one tutoring, project- and design-based classrooms, and computer-based learning environments, including ITSs.

The Evolved (Current) Conception of Scaffolding

A key element of the scaffolding process is that a student is able to achieve goals with assistance that he or she would not be able to do on his or her own (Wood et al., 1976). Traditionally, scaffolding includes an ongoing assessment of the learner's current state, support that is calibrated for the learner's state, fading of assistance, and transfer of responsibility to the learner (Puntambekar & Hubscher, 2005; van de Pol et al., 2010). These elements are consistent with the idea of scaffolding as a temporary structure that

will assist students in initial learning, and will then later be removed when they are able to achieve the goal without assistance (Lajoie, 2005).

Scaffolding was originally conceptualized for use with one-to-one human tutoring and instruction. However, as instruction has shifted to involve more complex environments, which include hypermedia, online learning, and ITSs, the practice of using scaffolding has changed (Puntambekar & Hubscher, 2005). There has been inconsistency in the definition of scaffolding in the recent research literature, with elements of the practice such as fading often missing. It has been suggested that in many cases what is referred to as scaffolding is actually guided instruction or support, where the learner is offered guidance and feedback, but does not receive critical elements of the scaffolding technique. There has recently been a move toward redefining and rethinking scaffolding in terms of both the original components and the new computer-based contexts that learning occurs in (Pea, 2004; Puntambekar & Hubscher, 2005; van de Pol et al., 2010; Yelland & Masters, 2007). This shift from person-based to computer-based instruction has created new issues and challenges to each of the critical characterizations of scaffolding: contingency (ongoing diagnosis and calibrated support), fading, and transfer of responsibility to the learner.

Changes in Scaffolding Characterizations

Ideally, in one-on-one tutoring, between a human tutor and a student, the tutor is able to observe the behavior and performance of the student. This ideal human tutor can intuitively notice shifts in the student's mood or performance, and then adjust the type of instruction accordingly. Further, the tutor can then use their own previous experience, knowledge of the student, and observations of the student's actions to help guide learning. The ideal individual human tutor can dynamically adjust to the situation, review material that they feel is not being grasped, and aid the learner in reaching their goals. The process of a human tutor successfully teaching and adjusting to an individual is a skill that is effortful, takes attention, and sometimes requires the tutor to respond in ways that are unique to the individual student as opposed to being consistent with the tutoring that others receive. In Wood et al. (1976), a script was designed to be used to provide specific feedback to all participants (3-, 4-, and 5-year-old children); however, it was found that 4 year olds presented behavior that required deviations from the tutoring script. This deviation from the script in order to support an individual's learning required the attention and judgment of the human tutor (Wood et al., 1976).

While an ideal human tutor can intuitively assess the learner's state, a computerized tutoring system needs to rely on input from surveys and sensors in order to determine state information. Careful consideration must go into selecting ways of accurately assessing and establishing the learner's state. The ITS can then use the learner's state to select what is expected to be appropriate guidance and feedback for the situation. Further, one-on-one tutoring may require building a rapport and trying different methods before finding the optimum guided learning path for the individual. In traditional scaffolding, the relationship between the adult expert and the student is very important. However, by shifting to computer-based systems, this relationship is no longer present. The relationship is now between the computerized system and the learner. However, there are ways to assist in making the interaction between the learner and the computer more similar to a human interaction. For instance, research has found that presenting material in a conversational manner, as opposed to formal, elicits better learning outcomes in multimedia environments (Moreno & Mayer, 2000, 2004). Conversational language may elicit social schemas that keep learners more engaged in the learning process and make them feel that the system has a social presence (Moreno & Mayer, 2004; Nass, Steuer & Tauber, 1994). By using this and similar techniques, the computer can engage the student in the lesson and assist in recreating the social aspect of scaffolding.

In scaffolding, there is the challenge of knowing when to fade and "remove" the scaffold to allow the individual to pursue the material on his or her own. This removal of the scaffold then transfers the

responsibility for learning to the student rather than the system (van de Pol et al., 2010). Knowing when to fade instruction is a challenge inherent in guided instruction and scaffolding with human tutors, but is even more of a concern with a computerized system. In computer-based learning, the calibration of difficulty level for the individual student is even more important. If the level of the work is too easy with the assistance, the learner may begin to lose interest in the system and not pursue further learning from it (Lajoie, 2005). Further, if the level of scaffolding and guidance is inappropriate or too hard for the individual, he or she may become discouraged and not want to continue working. An ideal human tutor will be able to pick up on these mood shifts and quickly readjust the lesson to reengage the learner. While designing a computer-based tutor that is sensitive to a student's emotion is a challenge, if done correctly, it can result in a system that may be more successful than a human tutor in identifying emotion shifts and reacting to them (D'Mello & Graesser, 2012).

Computer Based Learning Environments (CBLEs)

The definition of scaffolding has shifted through the years to also include more than simply one-to-one human tutoring. The term "scaffolding" has also more recently been used to include teacher to entire-class instruction, peer to peer instruction, and group-based project learning (Puntambekar & Hubscher, 2005). The definition has also specifically been re-conceptualized to include the addition of technology. Yelland and Masters (2007) proposed a new category of scaffolding, called *technical scaffolding*. They defined it generally in terms of using a computer as the medium for scaffolding, and the impact that it has on the instruction. However, we believe that it is important to further distinguish between types of computerized instruction, as they often range from highly student controlled (e.g., online college classes) to highly system controlled (e.g., ITSs). While these are all computer-based, the different characteristics they feature may impact student's learning, and necessitate differing scaffolding strategies. The literature includes references to a number of different computer-based learning types, which have been summarized and defined below:

Hypermedia

Hypermedia can include computerized lesson and materials that are presented to the individual on a computer. It can include videos, audio, and text (Azevedo, Cromley & Seibert, 2004). In some cases, it might be presented in addition to classroom instruction or completely on its own without specific instructions. Students may independently engage with a series of hypermedia environments, which can lead to difficulty in making sense of the connections between the material. However, it can be beneficial to design hypermedia environments with consistency and care to be coherent and understandable (McNamara & Shapiro, 2005).

Online Inquiry

It has become increasingly important that students, especially those in middle and high school, understand how to use the Internet to do research. As part of class assignments, students are often given a question that they need to answer through Internet searches. These assignments can occur in an environment where the students have access to the teacher or in the form of homework. Examples of scaffolding for online inquiry include 1) teacher-provided questions that enable the student to organize information from their web-searches and 2) providing students with software that offers a structure to break down the assignment into pieces and organize the main points of articles (Quintana, Zhang & Krajcik, 2005; Zhang & Quintana, 2012).

Online and Web-Based Instruction

Many colleges of today offer online courses. While these courses have an instructor that can provide guidance through emails and attending office hours, the instruction is primarily provided online. The instructor provides a framework for the students, creates computer-based lessons, and sets up deadlines for assignments. These materials are generally accessed through a Learning Management System (LMS). The social aspect of this instruction has been largely removed, which requires the students to use their own time-management and learning strategies. However, students may not always have the appropriate learning strategies to successfully regulate their online learning (Graesser & McNamara, 2010; Lim, 2004). Scaffolding can occur in online and web-based instruction through materials provided by the instructor and through the assignments that are given, which can assist in leading the student through the material. Further, the tools within the LMS, such as discussion boards, chatrooms, and resource links, can be used to engage, scaffold, and guide students through the learning process (Dabbagh, 2003; Dabbagh & Kitsantas, 2013).

Static Scaffolding in Computer-Based Learning

A critical component of scaffolding is contingency, which consists of monitoring the learner's state and adapting the instruction in a manner that is sensitive to the learner's state. However, often tutorials and computer-based instructional programs are developed with static scaffolding, which guides the learner through the material through unchanging standard feedback or by simply structuring the student's interaction (Molenaar, Roda, van Boxtel & Sleegers, 2012). This feedback and instruction may provide support and encourage the students to use metacognitive best practices; however, without adapting to the specific individual it is not consistent with the traditional definition of scaffolding. In static computer-based learning, the student may be provided with computer-based lessons, which are worked through in specific sequences and have guidance. However, each and every individual who works through the system receives the same feedback at the same point in the lesson (Molenaar et al., 2012).

Dynamic Scaffolding in Computer-Based Learning

Dynamic scaffolding occurs when the performance and state of the student is continuously assessed, and then the materials are adjusted as a result (Molenaar et al., 2012). This adaptation to the student's current state and performance makes this type of system consistent with the initial conceptualization of scaffolding. These are also features that are generally included in ITSs. While previous scaffolding literature has offered divisions of cognitive, metacognitive, affective, and technical scaffolding (van de Pol et al., 2010; Yelland & Masters, 2007), the increasing use of technology has both added new categories of scaffolding and shifted the importance of others. It is now vital to consider what type of instruction is being given – in person or computer-based. Further, as distinguished above, it is important to think about the benefits and limitations of different types of computer-based learning (e.g., hypermedia, online learning, ITS). In situations where the student primarily interacts with the computerized system and does not receive in-person instruction, motivation, self-regulation, and metacognition increase in importance.

Scaffolding in Computer-Based Learning

Shifting toward computerized learning puts more responsibility on the student than in traditional scaffolding. Computer-based learning is sometimes experienced without instructor guidance, so the student is tasked with regulating his or her own learning. Students have to make choices about what order to learn information, how long to spend on information, and how to manage their time wisely. It has often been found that students have difficulty successfully regulating their learning in hypermedia environments

(Azevedo & Hadwin, 2005); however, scaffolding within computer-based systems may improve their learning and performance. As a result, the current literature review focuses on self-regulatory and meta-cognitive scaffolds, which are of particular help to students who have to manage their own time and learning in computer-based systems.

Scaffolding for Self-Regulated Learning

In the current conception of scaffolding, process-oriented scaffolding has increased in importance, especially for CBLEs. In CBLEs, learners must be able to regulate their own learning with regard to decision making, time on tasks, navigation of instructional material, and maintaining engagement; however, not all learners have this ability inherently. Moreover, learners tend to have difficulty self-regulating and producing learning gains in hypermedia environments that teach complex topics (Lajoie & Azevedo, 2006). There are four types of scaffolds commonly accepted and implemented in hypermedia environments: 1) *conceptual* scaffolding uses hints and pumps for knowledge to use for problem solving; 2) *metacognitive* scaffolding uses human or nonhuman agents to provide help for task-related tasks; 3) *procedural* scaffolding provides guidance on how to use resources or perform tasks; and 4) *strategic* scaffolding exposes students to different solution paths and different techniques for problem solving (Hannafin, Land & Oliver, 1999). While it is common for CBLEs to these different types of scaffolds, there needs to be more empirical research evaluating the effectiveness of such scaffolds on learners' self-regulated learning. The notion is that learners must use several self-regulatory processes in order to effectively navigate and learn in CBLEs.

A self-regulated learner is able to maintain active engagement with the environment and determine which representations of material are most useful based on their self-knowledge, beliefs, motivation, domain and strategic knowledge, and task goals and definitions. A significant amount of research has been dedicated to scaffolding to promote SRL processes in hypermedia environments. Researchers are looking to SRL models for increasing our understanding of the particular SRL processes that are connected to learning in CBLEs, how the learner characteristics impact SRL, and how SRL can best be supported in such environments. In this section, we provide a synopsis of these studies. The framework of SRL outlines four areas of regulatory activity: cognition (goal-setting, employing and monitoring of cognitive strategies); motivation (self-efficacy, value of the task, interest); behavior (help-seeking, maintenance and monitoring, time use); and context (evaluation and monitoring of changed tasks) (Pintrich, 2000). Each of these areas is encompassed within each of the four phases of SRL: 1) planning, 2) self-monitoring, 3) control, and 4) evaluation. From a social cognitive perspective, SRL refers to the degree to which a learner is able to become metacognitively, motivationally, and behaviorally active participants in their own learning process (Zimmerman, 2000).

A major segment of related literature pertains to the effectiveness of scaffolds in hypermedia environments that foster SRL. The following studies have demonstrated the effectiveness of adaptive human tutoring (scaffolding) in facilitating learning with hypermedia. The ultimate goal is to develop models that can inform the future design of computer-based scaffolds in such environments. Azevedo, Cromley, and Seibert (2004) evaluated the impact of different scaffolding instructional interventions in facilitating learners' shift to more sophisticated mental models as an indication of both performance and process data. Fifty-one undergraduates were trained to use a hypermedia environment to learn about the circulatory systems. The participants were assigned to three different scaffolding conditions (adaptive scaffolding, fixed scaffolding, and no scaffolding). In the fixed scaffolding condition, learners were provided 10 domain-specific subgoals as fixed scaffolds to guide their learning. In the adaptive scaffolding condition, the learners were provided the same 10 subgoals as in the fixed scaffolding condition; however, a human tutor provided the individual support for the student based on ongoing diagnosis of the learner's current level of understanding. The participants' of the adaptive scaffolding conditions were found to enhance

their mental models significantly more than the other two conditions. Participants of this condition was also better able to regulate their learning by stimulating prior knowledge, using several strategies to monitor their emerging understandings, and engaging in adaptive help-seeking (Azevedo et al., 2004).

In a follow-on study, Azevedo, Cromely, Winters, Moos, and Greene (2005) also evaluated the same three scaffolding conditions (adaptive scaffolding, fixed scaffolding, or no scaffolding) on 111 adolescents learning about the circulatory system with a hypermedia environment. For this study, they used think-aloud protocols to help examine the impact of the condition on the student learning. This study also found that the adaptive scaffolding condition enhanced learners' mental models significantly more than the other conditions; however, learners in both the adaptive scaffolding and no scaffolding conditions gained significantly more declarative knowledge than learners in the fixed condition. While the adaptive scaffolding condition participants regulated their learning by the same methods found in the previous study, those in the no scaffolding condition used fewer effective strategies and those in the fixed scaffolding condition used processes that negatively impacted their learning (Azevedo, Cromley, Winters, Moos & Greene, 2005).

Moos and Azevedo (2008) investigated the impact of conceptual scaffolds on the planning and monitoring of SRL processes and self-efficacy. They collected self-reported data and think-aloud data from 37 undergraduates as they were interacting with a commercial hypermedia environment. The participants were assigned (randomly) to one of two experimental conditions: no scaffolding or conceptual scaffolding. According to the results, both conditions increased learners' self-efficacy toward the task and participants in the conceptual scaffolding condition used more SRL processes pertaining to planning on average than the no scaffolding condition (Moos & Azevedo, 2008).

Essentially, these researchers discovered that adaptive human tutoring is beneficial to learning and improving learners' SRL processes in hypermedia environments. Their findings had also produced an additional research need of not only understanding how students learn with hypermedia, but also how much students regulate their learning and how external regulating agents, such as human tutors, can facilitate learners' SRL. Azevedo, Moos, Green, Winters, and Cromley (2008) assessed how SRL and externally facilitated SRL (ERL) differentially impact adolescents' (N=128 middle and high school students) learning about the circulatory system with the use of hypermedia. Learners in the SRL condition regulated their own learning and learners in the ERL condition had a human tutor to facilitate their SRL. Both self-reported data and think-aloud data were collected. The results indicated that learners in the ERL condition gained significantly (statistically) more declarative knowledge and a higher number of participants produced a more advanced mental model as compared to the SRL condition (Azevedo, Moos, Greene, Winters & Cromley, 2008).

Azevedo, Cromely, Moos, Greene, and Winters (2011) conducted a more recent study on the effectiveness of three human scaffolding conditions in facilitating learners' learning about the domain (circulatory system) and the implementation of SRL processes using a hypermedia environment. The students (N=123) were randomly assigned to one of the three conditions: adaptive content and process scaffolding (ACPS), adaptive process scaffolding (APS), and no scaffolding (NS). In the APS condition, the human tutor provided process scaffolding, i.e., scaffolding learning by helping students perform key self-regulatory processes of planning, monitoring, and using different strategies. In the ACPS condition, the tutor provided content scaffolding, i.e., scaffolding about learners understanding of the domain content, in addition to process scaffolding. As expected, learners in the ACPS condition gained significantly (statistically) more declarative knowledge than the learners in the other conditions. They also used a small amount of regulatory processes, relied more on the tutor, and engaged in help-seeking behavior. The APS condition learners regulated their learning by using monitoring activities and learning strategies, but the learners in the no scaffolding condition were less effective at regulating their learning (Azevedo, Cromley, Moos, Greene & Winters, 2011).

The learning domain of these studies was the circulatory system, although similar studies have been conducted in other domains. Kramarski and Gutman (2005) conducted a study to compare two mathematical e-learning environment conditions: 1) e-learning with self-metacognitive questioning (EL+IMP) and 2) e-learning without support of self-regulation (EL). Sixty-five junior-high students were randomly assigned to one of the two environments to learn mathematical problem solving. The results indicated that the participants in the EL+IMP condition significantly outperformed the EL participants in problem-solving procedural and transfer tasks and in using self-monitoring strategies during problem solving (Kramarski & Gutman, 2005). Greene, Bolick, and Robertson (2010) evaluated how high-school students (*N*=40) used a hypermedia learning environment to obtain declarative knowledge of a historical event and historical thinking skills. They found that students are most engaged in the strategy use of SRL processes although their uses of planning SRL processes were most predictive of learning. The researchers suggest the scaffolding planning skills is useful for facilitating learners' use of computers as tools for learning (Greene, Bolick & Robertson, 2010). Unfortunately, the adaptive human scaffolding was not a part of these two studies.

A study conducted by Dabbagh and Kitsantas (2005) supports scaffolding for SRL through the use of web-based pedagogical tools, such as collaborative and communications tools, content creation, and delivery tools. They investigated how these tools supported the SRL processes of goal setting and self-monitoring as well as student's completion of assignments. Sixty-five students of three online courses were surveyed on their motivated strategies for learning, web-supported self-regulation, and perceptions of the usefulness of the web-based pedagogical tools they were using. They reported that different web-based pedagogical tools support SRL in different ways, and such tools are very effective for activating student's use of SRL processes essential to supporting specific types of learning tasks required for completion of course assignments (Dabbagh & Kitsantas, 2005).

Most of the studies pertaining to scaffolding and SRL in computer-based learning environments are predominantly conducted in open-ended, hypermedia environments. However, the notion of externally regulating learning is an important future direction of ITS research. Hadwin, Wozney, and Pontin (2005) suggested that this type of research is useful for the design of pedagogical agents that can support learning (Hadwin, Wozney & Pontin, 2005). Pedagogical agents are visual tutor representations that are often embedded within computer-based learning environments to establish a personal relationship and emotional connection with the learner. Biswas, Schwarts, Leelawon, Vye (2005) developed a teachable agent, called Betty's Brain, in which students learn by teaching Betty the necessary concepts pertaining to the domain of river ecosystems. In this study, Betty's Brain combines teaching by learning techniques and self-regulated mentoring to promote deep learning. The researchers focused on the components defining the learner-teacher interactions as well as the value of adding self-regulation hints to the mentor agent. The study compared three versions of the system to evaluate effectiveness: 1) a system in which the learner was tutored by the pedagogical agent (ITS); 2) a learning by teaching system (LBT) – students are taught with a basic version of Betty, then receive help from a mentor agent; and 3) a learn by teaching system where information on how to become better teachers and learners (SRL). Results indicated that self-regulation strategies to Betty and the mentor agent was the most effective condition and better prepared students to learn new concepts (Biswas, Schwartz, Leeawong & Vye, 2005). Since this study, the developers of Betty's Brain has continued their research on modeling and measuring self-regulated learning in teachable agent environments (Kinnebrew & Biswas, 2011; Kinnebrew, Biswas, Sulcer & Taylor, 2013; Roscoe, Segedy, Sulcer, Jeong & Biswas, 2013)

Another example of an ITS developed for SRL scaffolding is called MetaTutor. MetaTutor is a learning tool that teaches and trains students to self-regulate as they learn about several complex human body systems. It is also a research tool that collects data on students' cognitive, metacognitive, affective, and motivational processes produced during learning. Previous studies of MetaTutor have examined the effectiveness of SRL training versus no training on learners' ability to deploy SRL processes and found

SRL training to significantly outperform the control group (Azevedo, Witherspoon, Graesser, McNamara, Chauncey, Siler, Cai, Rus & Lintean, 2009). These studies have also found prior domain knowledge to significantly relate to how students' self-regulate their learning (Moos & Azevedo, 2008).

The studies presented above demonstrate the importance and effectiveness of scaffolding for self-regulated learning in computer-based learning environments. Unfortunately, there are few studies that evaluate the effectiveness of SRL techniques in ITSs; consequently, much more research is needed in the area of SRL scaffolding in ITSs. One of the benefits of scaffolding self-regulated learning is that it addresses the cognitive, affective, and motivational elements of learning. These elements are not only influential to the classification of learner states and performance, but are also essential to the appropriate implementation of instructional strategies. A primary process of SRL involves metacognition. The second major body of research pertaining to scaffolding in computer-based learning environments is on metacognitive scaffolding. The next section of this literature review provides more details.

Metacognitive Scaffolding

Metacognition is traditionally thought of as an individual's understanding of his or her own cognition (Flavell, 1979). In other words, it is thinking about one's own thinking. Metacognitive abilities become particularly important when one is asked to engage with computerized and online learning. An area of particular interest is online inquiry, which requires the learner to conduct a search of websites, gather information, make judgments about the utility of that information, and monitor their own learning (Quintana et al., 2005; Zhang & Quintana, 2012). Metacognition can be further broken down to include knowledge of one's own strengths and weaknesses as a learner, knowledge of the task to be accomplished, and knowledge of strategies that can be used to achieve one's own learning goals (Quintana et al., 2005). There are a number of different measures that have been developed to assess an individual's metacognitive abilities in general, as well as in context of specific skills such as reading (Moore, Zabrucky & Commander, 1997; Schraw & Dennison, 1994). These scales can then be used to classify individuals into high and low metacognitive ability categories. Those who have high metacognition scores often use strategies that assist them in learning and have a better understanding of what they know than those who have lower scores. However, these skills can be scaffolded and used to improve the learning of those who do not initially have high metacognitive abilities.

In online and computerized environments, the student is required to engage in the following metacognitive behaviors: 1) understanding their learning task and plan accordingly for it, 2) monitoring their current progress on that task and making adjustments in their strategies accordingly, and 3) reflecting on their learning experience (Quintana et al., 2005). There has been evidence that without scaffolding students have trouble regulating their learning in CBLEs. However, research has begun to examine the impact of providing scaffolding that assists the students in performing metacognitive behaviors, which can lead to improved learning outcomes (Azevedo & Hadwin, 2005; Azevedo & Jacobson, 2008).

In-class learning is guided by a teacher, who provides specific instructions, lessons, and assignments to students. The teacher offers feedback and opportunities for students to ask for help when they need it. By shifting instruction to computerized and online learning environments, the responsibility of pacing the material shifts from the teacher to the student. Students often times do not have highly developed methods for successful learning (Lim, 2004). If they are novices, or have no background in the topic area, they tend to perform worse than high prior knowledge individuals because they may not have context for the material or know where to start (Moos, 2014; Shapiro, 2004). One of the main ideas behind scaffolding is that with the help and guidance of a more knowledgeable individual, learners are able to achieve goals that they would not be able to do on their own. This idea can then be extended to guided learning and computer-based learning. In computerized environments, providing a metacognitive framework that is

composed of prompts, questions, activities, and/or feedback, students will have a more productive learning experience (Graesser, D'Mello & Person, 2009; Wiley, Goldman & Graesser, 2009) The goal of metacognitive scaffolding is for students to learn computerized material in a way that enhances their retention of it, and also provides them with strategies that they can mimic and successfully use in other learning environments.

While it has been determined that unscaffolded learners do not perform well in hypermedia environments, little research has examined the types of scaffolds that are beneficial in computer-based environments (Azevedo et al., 2004). Research into computerized scaffolding has ranged from fully computer-based to teacher-enhanced scaffolding. In addition, the focus of scaffolding also varies, with some systems choosing to scaffold the specific topic or lesson material (e.g., algebra), whereas others choose to scaffold specific processes of learning, such as metacognition (Azevedo & Jacobson, 2008). The studies below emphasize the vast differences in approaches to metacognitive scaffolding in computer-based environments. These approaches range from using hypermedia environments (text, hyperlink, and video-based) to structured online systems to guide inquiry to ITSs.

Wu and Pederson (2011) examined the impact of using both computer- and teacher-based scaffolds in computer-based science learning. They point out that, in general, the literature has focused on either teacher-based support or computer-based scaffolding, not the integration of both. Middle school students ($N = 142$) engaged with a virtual learning environment that explained how volcanoes worked. Participants completed a series of computerized science inquiry tasks. Scaffolding was provided in 4 of 5 of the tasks that they completed. The type of scaffolding (continuous and faded) was varied, as was the timing of the appearance of teacher-based metacognitive scaffolding (early or late). The computer-based scaffolding offered hints and a framework for the assignment; the teacher-based scaffolding was in the form of questions and inquiry tasks that required the student to reflect and self-explain. They found that continuous computer-based scaffolding was better than faded scaffolding in assisting students with completing the tasks. There were no significant differences in learning/performance found in regard to the timing of teacher provided metacognitive scaffolding. However, in general, students felt that the assistance of the teacher and the questioning provided by them was useful (Wu & Pederson, 2011).

As computers have become a bigger part of the classroom environment, there has been an increase in the requirement for students to use the Internet as an educational research. One common practice is engaging in online inquiry, which requires the student to answer questions by finding materials on the Internet, assessing them, and synthesizing the information (Quintana et al., 2005; Zhang & Quintana, 2012). Zhang and Quintana (2012) created a computer-based tool called *IdeaKeeper* that provided scaffolding for online inquiry. Middle school students in groups of 2 ($N = 16$) were asked to generate questions, and sub-questions, which they were to answer using web-searches/online inquiry. One group interacted with the *IdeaKeeper* tool, which prompted them to answer questions and provide information for their visited websites, which facilitated their understanding of the material. An additional group only used Google and were asked to take their own notes. The students in the scaffolded condition were more focused on the websites that they encountered and retained more useful information from them than those in the Google only condition. The computer-based tool assisted the students in attending to useful information and structuring their learning with metacognitive best practices.

Rather than simply being an independent or pair activity, online and web-based inquiry can also be used for group projects. Raes, Schellens, De Wever and Vanderhoven (2012) examined the impact of scaffolding on a web-based collaborative project. Their study specifically examined domain-knowledge and metacognitive awareness after high-school-aged Flemish students ($N = 347$) completed a four-week computer-based inquiry project. Students engaged with the *Web-Based Inquiry Science Environment* (Slotta & Linn, 2009), which was a stable online learning environment. The students were tasked with answering a question, seeking information about it, finding information, then combining and synthesizing

the information to answer the question. The student groups were divided into four conditions: technology-enhanced, teacher-enhanced, both technology- and teacher-enhanced, and control (no scaffolds). All conditions had improvement in their domain-specific knowledge. Technology-based scaffolding improved metacognitive awareness. Further, the knowledge level of the individual had an impact, with those who were low prior knowledge receiving more benefits from teacher-based interventions than those who were high prior knowledge (Rae, Schellens, De Wever & Vanderhoven, 2012).

Metacognitive and reflective prompts have been used to scaffold students to design an experiment in a computer-based environment recommendations providing for future scaffolding research and ITSs (Morgan & Brooks, 2012). Morgan and Brooks (2012) used a computer program, which scaffolded high school chemistry students ($N = 102$) through developing a research study using a backwards design. The backwards design was meant to lessen cognitive load, as it was more consistent with the order that experiments are usually developed in, rather than forcing students into an order that did not make sense. There were four conditions, such that students received either a backwards design process, or a student-based one, and additionally received either reflective prompt scaffolding or not. While the backwards design was found to be successful, there were mixed results with the scaffolding. The lab report performance of those who were higher level students were not impacted by the scaffolding, whereas lower level students were hurt by the metacognitive prompts. The authors believed that the increase in cognitive load caused by the metacognitive reflection reduced the performance of the lower level students.

Yildiz-Feyzioglu, Akpinar, and Tatar (2013) examined students' metacognitive knowledge in a technology-enhanced learning environment when metacognitive prompts were used. They performed a descriptive case study of Turkish seventh grade students ($N = 3$), who were engaging with a computerized unit about electricity. Metacognitive prompts were used as an instructional tool in the learning environment. The students were encouraged to plan, set goals, monitor their progress, as well as reflect on and evaluate their learning. Two out of three of the students demonstrated an increase in metacognitive knowledge as a result of the prompts. The metacognitive scaffolding and prompts appear to have increased their confidence in the topic and lead to improved understanding of the material (Yildiz-Feyzioglu, Akpinar & Tatar, 2013).

Moos (2014) examined the relationship between motivation and the use of metacognitive strategies in hypermedia learning. The researchers had undergraduate students ($N = 85$) complete an essay about what they knew about the circulatory system to gauge prior knowledge. Afterward, the participants engaged with a hypermedia encyclopedia environment that included text, illustrations, hyperlinks, diagrams, and a video about how the circulatory system works. During the learning phase they were asked to "think aloud" and explain their thoughts and what they were doing at the time. The audio and video recordings of the learning process were coded and examined to determine the number of metacognitive processes that were engaged in during the "think aloud." Both self-efficacy and extrinsic motivation were found to be predictors of using metacognitive processes and monitoring their own understanding. This study showed that there are individual characteristics that may be related to the spontaneous use of metacognitive strategies in computerized environments (Moos, 2014).

The use of immediate feedback as a metacognitive scaffold has been examined within a medical ITS (El Saadawi et al., 2010). In the past, immediate feedback has been criticized, as it may prevent the individual from successfully developing metacognitive strategies. However, in certain domains, particularly in the medical field, it is extremely valuable in training individuals for the tasks they will be performing. Participants were medical residents ($N = 23$) who received training with the *SlideTutor* ITS. The first day of the training primarily consisted of training and assessments, and the second day included a metacognitive intervention for one group and no intervention for the control group. In addition, participants in both groups engaged in a faded feedback condition, which involved gradually removing feedback over time. The metacognitive interventions included activities that encouraged students to assess what they know,

and reflect on their own performance. Immediate feedback was found to lead to increased learning gains; however, the removal of the feedback did lead to negative effects. The metacognitive scaffolds did have a positive impact on the ability of students to judge if they were right or wrong, but did not affect performance on the assessments given. One explanation provided by the researchers is that the scaffolding only occurred over a two-day period, and gains may require more time and experience with the metacognitive scaffolding.

Molenaar, Chiu, Sleegers and van Boxtel (2011) used embodied computerized agents to administer scaffolding to elementary school students ($N = 54$) working in groups of three to complete an assignment. There were three conditions: control, structuring metacognitive scaffolding, and problematizing metacognitive scaffolding. They were interested in examining how these different types of metacognitive scaffolds would affect the student's metacognitive and domain knowledge. The students engaged in the *Ontdeknet* e-learning environment in which they had access to an expert. During completion of their assignment (to learn about a country and write a paper about whether they would like to live abroad) those in the structuring metacognitive scaffolding group received scaffolds that guided them through the assignment, gave them examples, provided support, and encouraged the students to plan. In the problematizing metacognitive scaffolding group, the students were asked questions that would require them to think about their activities and explain what they were doing. Both types of metacognitive scaffolds led to increased metacognitive knowledge. Problematizing scaffolds led to students demonstrating higher domain knowledge than those were in the structuring and baseline conditions. The researchers suggested that by providing questions in the problematizing condition, it encouraged the students to engage in metacognitive activities with their group. Therefore, different metacognitive strategies can lead to different outcomes (Molenaar, Chiu, Sleegers & van Boxtel, 2011).

The above studies demonstrate the wide range of approaches to metacognitive scaffolding. The computerized learning environments and level of interactivity they provide varied greatly, from full ITSs (El Saadawi et al., 2010) to hypermedia encyclopedias (Moos, 2014) to online inquiry web searches (Zhang & Quintana, 2012). Further, the source of the scaffolding also varied from being provided by teachers (Wu & Pederson, 2011) to being provided a list of structured questions (Zhang & Quintana, 2012) to being given by an avatar (Molenaar et al., 2011). This variety highlights the need to begin differentiating between the sources of metacognitive scaffolding, and the types of computerized environments and metacognitive strategies that they are successful in. The interactivity level of the CBLE may have an impact on the success of the scaffolding because the characteristics of the systems are very different.

Conclusions/Recommendations for GIFT

In sum, the scaffolding metaphor in its definition, conceptualization, and implementation has changed in the transition from traditional classroom-based and one-to-one tutoring situations to the progressive use of computer-based learning environments. The studies presented in this literature review chapter indicate that SRL and metacognitive scaffolding in CBLEs is critical for fostering learning. This process requires a complex interplay among learner characteristics, systems features, scaffolds, and learning processes, which can be difficult to control for experimentation and analysis. Consequently, the measurement and analysis of scaffolding is still in its infancy as there is no standardization for evaluating the effectiveness of scaffolds in CBLEs, especially for ITSs. A majority of the articles call for the need for more research in this regard. One of the benefits of GIFT is that it's designed to conduct comparisons of instructional scaffolding strategies for effectiveness, which would be beneficial toward solidifying standardized techniques.

The literature on the effectiveness of affective scaffolding in one-to-one human tutoring and classroom environments is scarce and is limited among the ITS literature. This shortage is expected as affective

computing research is still in its early stages; however, as mentioned previously, cognitive, conative, and affective factors are interconnected when it comes to the dimension of learning. A recommendation for GIFT is to conduct comparison experiments assessing the impact and effectiveness of these different types of scaffolding on learning outcomes. GIFT can even use pedagogical agents to facilitate this process. Additionally, technical scaffolding, a term coined by Yelland and Masters (2007), has also been found to be beneficial to learning, and may be an element the GIFT could also consider investigating. Unlike most scaffolding studies, Yelland and Masters (2007) understood that the learning environment itself could be an influential mediator for affecting student learning. Therefore, technical scaffolding should be considered in all future ITS research. In addition, it is important to start examining the different types of CBLEs that are used and the different influences they may have on student learning.

Another next step for ITS researchers and GIFT future development is to conduct more research on the fading and transfer of responsibility processes of scaffolding. As mentioned previously, fading is often the missing element implemented in current scaffolding strategies. GIFT can facilitate fading beyond its current implemented in its four-level assessments (i.e., at expectation, above expectation, below-expectation, unknown). As a recommendation, GIFT could use another mechanism to monitor the transition between below-expectation and at-expectation for implementing fading and the transfer of responsibility.

References

Azevedo, R., Cromley, J., Moos, D., Greene, J. & Winters, F. (2011). Adaptive Content and Process Scaffolding: A key to facilitating students' self-regulated learning with hypermedia. *Psychological Test and Assessment Modeling, 53*(1), 106-140.

Azevedo, R., Cromley, J. & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology, 29*, 344-370.

Azevedo, R., Cromley, J., Winters, F., Moos, D. & Greene, J. (2005). Adaptive human scaffolding facilitates adolescent's self-regulated learning with hypermedia. *Instructional Science, 33*(5-6), 381-412.

Azevedo, R. & Hadwin, A. (2005). Scaffolding self-regulated learning and metacognition -- Implications for the design of computer-based scaffolds. *Instructional Science, 33*, 367-379.

Azevedo, R. & Jacobson, M. (2008). Advances in scaffolding learning with hypertext and hypermedia: a summary and critical analysis. *Educational Technology Research Development, 56*, 93-100.

Azevedo, R., Moos, D., Greene, J., Winters, F. & Cromley, J. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia. *Educational Technology Research Development, 56*, 45-72.

Azevedo, R., Witherspoon, A., Graesser, A., McNamara, D., Chauncey, A., Siler, E., Cai, Z., Rus, V. & Lintean, M. (2009, July 9-10). MetaTutor: Analyzing Self-Regulated Learning in a Tutoring System for Biology. In *Proceedings of the 14th International Conference Artificial Intelligence In Education*, Brighton, UK.

Biswas, G., Schwartz, D., Leeawong, K. & Vye, N. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intleligence, 19*, 363-392.

Dabbagh, N. (2003). Scaffolding: An important teacher competency in online learning. *TechTrends, 47*(2), 39-44.

Dabbagh, N. & Kitsantas, A. (2005). Using web-based pedagogical tools as scaffolds for self-regulated learning. *Instructional Science, 33*, 513-540.

Dabbagh, N. & Kitsantas, A. (2013). Using Learning Management Systems as Metacognitive Tools to Support Self-Regulation in Higher Education Contexts. In *International Handbook of Metacognition and Learning Technologies* (pp. 197-211). Springer New York.

D'Mello, S., Lehman, B. & Graesser, A. C., (2011). A Motivationally Supportive Affect-Sensitive AutoTutor. In R. Calvo & S. D'Mello (Eds.). *New Perspectives on Affect and Learning Technologies* (pp. 113-126). New York: Springer.

D'Mello, S. K. & Graesser, A. C. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems*, *2*(4), 23: 1-38.

El Saadawi, G., Azevedo, R., Castine, M., Payne, V., Medvedeva, O., Tseytlin, E. et al. (2010). Factors affecting feeling-of-knowing in a medical intelligent tutoring system: the role of immediate feedback as a metacognitive scaffold. *Advances in Health Science Education, 15*, 9-30.

Flavell, J. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911.

Kinnebrew, J.S. and G. Biswas. Modeling and Measuring Self-Regulated Learning in Teachable Agent Environments," Journal of e-Learning and Knowledge Society, English Edition, vol. 7, no. 2, pp. 19-35, 2011.

Kinnebrew, J.S., Biswas, G., Sulcer, B. & Taylor, R.S. (2013). Investigating Self-Regulated Learning in Teachable Agent Environments. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies*, pp. 451-470. New York: Springer.

Graesser, A. C., D'Mello, S. & Person, N. K. (2009). Metaknowledge in tutoring. In D. Hacker, J. Donlosky & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 361-382). New York: Taylor & Francis.

Graesser, A.C. & McNamara, D.S. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist, 45,* 234-244.

Greene, J., Bolick, C. & Robertson, J. (2010). Fostering historical knowledge and thinking skills using hypermedia learning environment: The role of self-regulated learning. *Computers & Education, 54*, 230-243.

Hadwin, A., Wozney, L. & Pontin, O. (2005). Scaffolding the appropriation of self-regulatory activity: A socio-cultural analysis of changes in student-teacher discourse about a graduate research portfolio. *Instructional Science, 33*(5-6), 413-450.

Hannafin, M., Land, S. & Oliver, K. (1999). Open Learning Environments: Foundations, Methods, and Models. In C. M. Reigeluth (Ed.), *Instructional design theories and models* (Vol. 2, pp. 115-140).

Kramarski, B. & Gutman, M. (2005). How can self-regulated learning be supported in mathematical e-learning environments? *Journal of Computer Assisted Learning, 22*, 24-33.

Lajoie, S. (2005). Extending the Scaffolding Metaphor. *Instructional Science, 33*, 541-557.

Lajoie, S. & Azevedo, R. (2006). Teaching and learning in technology-rich environments. In P. A. Alexander & P. H. Winnne (Eds.), *Handbook of educational psychology* (pp. 803-821). Mahwah, NJ: Erlbaum.

Lim, C. (2004). Engaging learning in online learning environments. *TechTrends, 48*(4), 16-23.

McNamara, D. & Shapiro, A. (2005). Multimedia and hypermedia solutions to promoting metacognitive engagement, coherence, and learning. *Journal of Educational Computing Research, 33*(1), 1-29.

Molenaar, I., Chiu, M., Sleegers, P. & van Boxtel, C. (2011). Scaffolding of small groups' metacognitive activities with an avatar. *Computer-Supported Collaborative Learning, 6*, 601-624.

Molenaar, I., Roda, C., van Boxtel, C. & Sleegers, P. (2012). Dynamic scaffolding of socially regulated learning in a computer-based learning enviornment. *Computers & Education, 59*, 515-523.

Moore, D., Zabrucky, K. & Commander, N. (1997). Validation of the metacomprehension scale. *Contemporary Educational Psychology, 22*, 457-471.

Moos, D. (2014). Setting the stage for the metacognition during hypermedia learning: What motivation constructs matter? *Computers & Education, 70*, 128-137.

Moos, D. & Azevedo, R. (2008). Monitoring, planning, and self-efficacy with hypermedia: The impact of conceptual scaffolds. *Computers in Human Behavior, 24*, 1686-1706.

Moos, D. & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology*, 33(2), 270-298.

Moreno, R. & Mayer, R. (2000). Engaging students in active learning: The case of personalized multimedia messages. *Journal of Educational Psychology, 92*(4), 724-733.

Moreno, R. & Mayer, R. (2004). Personalized messages that promote science learning in virtual environments. *Journal of Educational Psychology, 96*(1), 165-173.

Morgan, K. & Brooks, D. (2012). Investigating a method of scaffolding student-designed experiments. *Journal of Science Education and Technology, 21*, 513-522.

Nass, C., Steuer, J. & Tauber, E. (1994). *Computers are social actors.* Paper presented at the SIGCHI Conference on Human Factors in Computer Systems, Boston, MA.

Pea, R. (2004). The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. *The Journal of the Learning Sciences, 13*(3), 423-451.

Pintrich, P. R. (2000). The role of goal-orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451-502). San Diego, CA: Academic.

Puntambekar, S. & Hubscher, R. (2005). Tools for Scaffolding Students in a Complex Learning Environment: What Have We Gained and What Have We Missed? *Educational Psychologist, 40*(1), 1-12.

Quintana, C., Zhang, M. & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist, 40*(4), 235-244.

Rae, A., Schellens, T., De Wever, B. & Vanderhoven, E. (2012). Scaffolding information problem solving in web-based collaborative inquiry learning. *Computers & Education, 59*, 82-94.

Roscoe, R.D., Segedy, J.R., Sulcer, B., Jeong, H. & Biswas, G. Shallow Strategy Development in a Teachable Agent Environment Designed to Support Self-Regulated Learning," *Computers and Education*, vol. 62, pp. 286-297, 2013.

Rosenshine, B. & Meister, C. (1992). The use of scaffolds for teaching higher-level cognitive strategies. *Educational Leadership, 49*(7), 26-33.

Schraw, G. & Dennison, R. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460-475.

Shapiro, A. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41*(1), 159-189.

Slotta, J. & Linn, M. (2009). *WISE science, web-based inquiry in the classroom*. New York: Teachers College Press.

Stone, C. A. (1993). What is missing in the metaphor of Scaffolding. In E. A. Forman, N. M. Minick & C. A. Stone (Eds.), *Context for learning. Sociocultural dynamics in children's development* (pp. 169-183). New York: Oxford University Press.

Stone, C. A. (1998). The metaphor of scaffolding: Its utility for the field of learning disabilities. *Journal of Learning Disabilities, 31*, 344-364.

van de Pol, J., Volman, M. & Beishuizen, J. (2010). Scaffolding in Teacher-Student Interaction: A Decade of Research. *Educational Psychology Review, 22*, 271-296.

Vygotsky, L. S. (1978). *Mind in society—The development of higher psychological processes*. Cambridge: Harvard University Press.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K. & Hemmerich, J. A. (2009). Source Evaluation, Comprehension, and Learning in Internet Science Inquiry Tasks. *American Educational Research Journal, 46*, 1060-1106.

Wood, D., Bruner, J. S. & Ross, G. (1976). The role of tutuoring in problem solving. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 17*(2), 89-100.

Wu, H.-L. & Pederson, S. (2011). Integrating computer- and teacher-based scaffolds in science inquiry. *Computers & Education, 48*, 362-382.

Yelland, N. & Masters, J. (2007). Rethinking scaffolding in the information age. *Computers & Education, 48*, 262-382.

Yildiz-Feyzioglu, E., Akpinar, E. & Tatar, N. (2013). Monitoring students' goal setting and metacognitive knowledge in technology-enhanced learning with metacognitive prompts. *Computers in Human Behavior, 29*, 616-625.

Zhang, M. & Quintana, C. (2012). Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education, 58*, 181-196.

Zimmerman, B. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. Pintrich & M. Seidner (Eds.), *Self-Regulation: Theory, Research and Applications*. Orlando, FL: Academic Press.

CHAPTER 23 – Adaptive Multimedia Environments

Wolfgang Schnotz[1] and Whitney Cade[2]
[1]University of Koblenz-Landau; [2]University of Memphis

Introduction

Multimedia environments combine multiple forms of representations, such as texts, tables, pictures, and graphs, as well as multiple sensory modalities, such as the auditory modality (when listening to spoken text or sound), the visual modality (when reading text or looking at pictures or graphs), or other modalities (such as touch), into an integrated configuration of information delivery (Mayer, 2005). Multimedia environments do not necessarily require electronic information technologies. There can be printed books and blackboards instead of computer screens, as well as human voice instead of audio speakers. Computer-based multimedia environments have the potential to adapt to the information and comprehension needs of different groups of learners with different aims (Mayer, 2009).

In this chapter, we describe different approaches to the design of adaptive multimedia environments. We begin by pointing out the specific functions of different forms of representations as well as the specific functions of different sensory modalities for comprehension of multimedia messages and multimedia learning. We subsequently consider the possibilities of implementing adaptive constraints into a multimedia environment in a rule-based format, namely, in terms of production rules. We also perform a corresponding analysis within an alternative architecture, namely, a constrained network activation approach. Finally, we draw some general conclusions and make suggestions for the design of multimedia components in the GIFT framework.

Related Research

Forms of Representations in Multimedia Environments

The different components of multimedia environments, such as written or spoken text, realistic pictures, or graphs (such as line graphs or bar graphs), correspond to different forms of representations that serve different purposes (Ainsworth, 1999; Peterson, 1996). The primary texts that are expected to be read provide conceptual guidance, whereas realistic pictures and graphs (possibly combined with short explanatory texts) typically serve as external cognitive tools on demand for mental model construction, which are frequently not used as carefully as expected. The primary texts guide the reader's conceptual analysis by describing the subject matter. They are closely related to the conceptual (propositional) representation that can be further used for mental model construction. Realistic pictures are two-dimensional simulations of objects or scenarios from a specific perspective, whereas graphs are analog representations (models) of some facts that possibly include also abstract, imperceptible relationships. Realistic pictures and graphs typically serve as scaffolds for initial mental model construction first, but are afterwards more likely to be used as easily accessible external representations for model updates, if needed. In other words, initial mental model construction is primarily text-based, whereas further task-oriented elaboration of the mental model relies on a more detailed analysis of accompanying pictures (Hochpöchler et al., 2013).

Numerous studies have shown that students usually learn better from words and pictures than from words alone (Mayer, 2009). However, this effect is bound to specific conditions. Text comprehension and picture comprehension are different routes for constructing mental models, whereas picture comprehen-

sion becomes more important when learners are poor readers than when they are good readers. When learners have low prior knowledge, adding a picture as another source of information can enhance comprehension because it offers an additional route for mental model construction. Learners with high prior knowledge are frequently able to construct a mental model also without pictorial support.

Adding pictures to texts can be a double-edged sword. On the one hand, learners frequently process pictures only superficially, because they assume that a quick inspection is enough to grasp the meaning (Hannus & Hyönä, 1999; McDonald & Thornley, 2002; Mokros & Tinker, 1987; Weidenmann, 1989). If this is correct, instructional design has to adopt methods to ensure sufficiently deep picture comprehension (Bernard, 1990; Kulhavy, Lee & Caterino, 1985). Learners also need sufficient spatial cognitive skills to comprehend the picture (Mayer, 1997). However, it should be acknowledged that if a picture is added to a text, the text information becomes less important. Thus, the text is frequently processed less deeply than if the text had been processed without pictures (Schnotz & Bannert, 1999). Because learners with high prior knowledge frequently do not need both text and pictures as information sources, adding a picture to a written text possibly means adding redundant, unneeded information. Although one of the two information sources is not needed, the learner's eyes wander between the two sources causing a split in attention. The student invests time and effort into processing redundant information without any learning benefit. This negative effect is called the "redundancy effect" (Chandler & Sweller, 1996; Kalyuga, Chandler & Sweller, 2000). These considerations apply also to graphs, but one has to keep in mind that graphs require additional visual literacy guided by graphic schemata (Pinker, 1990).

Modalities in Multimedia Environments

Besides multiple forms of representations, multimedia messages make use of multiple sensory modalities, notably the visual and auditory modalities. Written text, pictures, and graphs are processed by the visual modality whereas spoken text and sound by the auditory modality. Regardless of the increasing similarity of reading and listening competence under the condition of adequate education, visual and auditory texts differ in terms of cognitive processing. If presented singularly (i.e., without accompanying pictures or graphs), visual text leads to better memory for micro-propositions (details) than auditory text, whereas the latter leads to better memory for macro-propositions and the gist of the text (Hildyard & Olson, 1978; Rubin, Hafer & Arata, 2000). On the one hand, visual text is usually stable, which provides more control of cognitive processing, because it allows re-reading difficult passages, whereas spoken text is transitory. On the other hand, auditory text allows the student to take full advantage of text-picture combinations by maximizing temporal contiguity of verbal and pictorial information in working memory and by minimizing the negative effects of split attention. These considerations have led to the assumption of a modality effect, suggesting that pictures should be combined with auditory rather than visual text (Mayer & Sims, 1994).

Although the modality effect seems to be a well-established result of empirical research, there is no straightforward answer to the question of where the effect comes from. The most prominent explanations are the avoidance of split attention if both modalities are used and cognitive overload if only visual working memory is involved (Mayer & Moreno, 1998; Moreno & Mayer, 1999). The combination of spoken text and pictures allows the student to make the most of the multimedia effect (i.e., the combination of text and pictures), but a modality effect is only to be expected if there is also a multimedia effect: If there is no multimedia effect, no modality effect is to be expected either. If pictorial support for mental model construction is not essential for the learner because the leaner has sufficient prior knowledge, there is no modality effect to be expected.

Individuals also do not have to look at a picture continuously when processing a multimedia message. A modality effect is more likely to occur when the learner had not seen the pictures before and the learner

needs to look at the picture in order to process the paragraph appropriately. One consequence of this is the need for split attention between text and picture. If, however, the picture has already been presented before with the preceding content-related paragraph, the picture is no longer new to the learner. In this case, it is possible that the learner already has the required pictorial information in working memory due to previous processing. Thus, there is no need any more to look at the picture while processing the corresponding picture-related paragraph. Complex pictures might be more difficult to hold in working memory so there is the need for more glances at the picture to update its representation in working memory. Furthermore, a complex picture needs more verbal guidance and explanation and, thus, results in longer and more complex picture-related paragraphs than a simple picture. Thus, longer picture-related paragraphs might be associated with a stronger need for split attention than simple pictures with shorter paragraphs.

In summary, a modality effect is to be expected to the extent that split attention is required between visual text and pictures. The need for split attention is especially high for picture-related paragraphs, which suggests that picture-related paragraphs should be presented generally in the auditory modality. This is especially true if pictures have not been seen before by the learner. Whether content-related paragraphs should also be presented in the auditory or visual modality depends on how much split attention is required. There is no reason to assume a strong need for split attention if the picture is relatively simple, if learners have relatively high prior knowledge, or if the density of semantic connections between text and pictures is relatively low. In these conditions, visual text could be more advantageous because it allows for better control of cognitive processing by the learner. Moreover, in no cases do individuals learn better from pictures accompanied by both spoken *and* written text.

Discussion

Rule-Based Multimedia Adaptation

Given the complexity of the interrelations described above, it is no surprise that there are no simple rules of thumb for designing and adapting multimedia environments. Design and adaptation of these environments involves multiple constraints satisfaction and requires a sufficiently deep understanding of the human cognitive system and its interaction with multimedia messages in order to formulate appropriate instructional strategies for adaptive multimedia environments.

Zone of Proximal Development

We consider instructional strategies as conditionalized procedural knowledge for the intentional guidance of individuals' activities to make them learn. These strategies are organized in hierarchies with superordinate and subordinate strategies, where the latter are sometimes called "tactics." One of the most general superordinate strategies in the design of adaptive multimedia environments is that any instruction should take place within Vygotsky's ZPD by demanding neither too little nor too much from the student (Vygotsky, 1978; Wood, Bruner & Ross, 1976). Negatively formulated, this implies that multimedia environments should not provide instructional aids that cannot yet be used by the learner or which are not needed any more. Positively formulated, it means that instructional aids should only be provided when learners need them and when they are able and willing to use them. Figure 1 demonstrates the relationship between the ZPD and the effectiveness of instructional aids. The upper panel shows the dependency of learners' performance on some task without aid (dashed line) and with some instructional aid (solid line). The task fits best to the ZPD of a learner at state L3, but is still acceptable for learners at state L2 or state L4. For a learner at state L1, the task is too difficult because there is no chance of solving the task even with a maximum of instructional aids: probability of success is 0%. For a learner at state L5, the task is

too easy, because he will perform the task successfully with a probability of 100%. The effectiveness the instructional aid is represented in the upper panel by the difference between the dashed line and the solid line, and is shown by the solid curve in the lower panel. Generally speaking, any learning task combined with some instructional aid can only be effective within some range of expertise, namely the ZPD. If an aid cannot be used yet or if it is not used any more, it will not be instructionally effective.
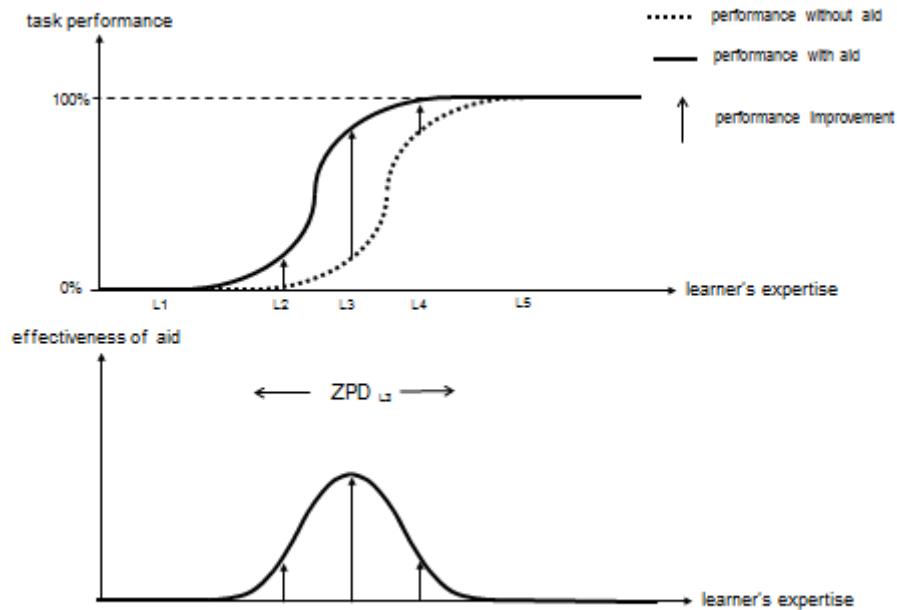
Figure 1. Relationship between the ZPD and the effectiveness of instructional aids.

Learner's *level of expertise* is determined by multiple factors, such as prior knowledge, reading literacy, visual literacy (especially regarding graph reading), spatial cognitive skills, and experience with handling computer-based multimedia learning environments. Similarly, the *difficulty of multimedia learning tasks* are affected by characteristics of texts, pictures, graphs, animations and the navigation demands of the environment and the abstractness of content. As for texts, these factors are, for example, familiarity and meaningfulness of words, complexity of syntax, text length, and different aspects text coherence, all of which can be determined automatically using a tool such as Coh-Metrix (Graesser, McNamara, Louwerse & Cai, 2004; McNamara, Graesser, McCarthy & Cai, 2014). Regarding pictures and graphs, these factors include graphical complexity in terms of graphical entities and relations, the number of depicted varia-bles, or the number and kind of cohesion devices between the picture/graph and the text (cf. Kirsch & Mosenthal (1990). For the navigation demands, relevant factors are the complexity of the animation space and the variety and operability of navigational tools. *Instructional aids* can be of multiple kinds, ranging from advance organizers, learning goals, verbal signaling, adding pictures or graphs, pictorial signaling, animations, illustrative or worked examples, and others.

Based on the relationships among the learner's expertise level, the learning task difficulty, and the task performance shown in Figure 1, it is possible to derive the relative instructional effectiveness of different instructional aids at different levels of expertise. Figure 2 shows the relative instructional effectiveness of different aids that can be derived from the effectiveness of different aids. According to the concept of ZPD mentioned above, instructional aid B is only more effective than alternative aids A and C within a

specific range of expertise. Outside this range, other aids (as, for example in this case, A and C) are more effective. Adaptive multimedia environments have to take these interdependencies between learner's level of expertise, difficulty of learning task, and conditional effectiveness of instructional aids into account in order to adequately decide which aid should be made available and which should be not.
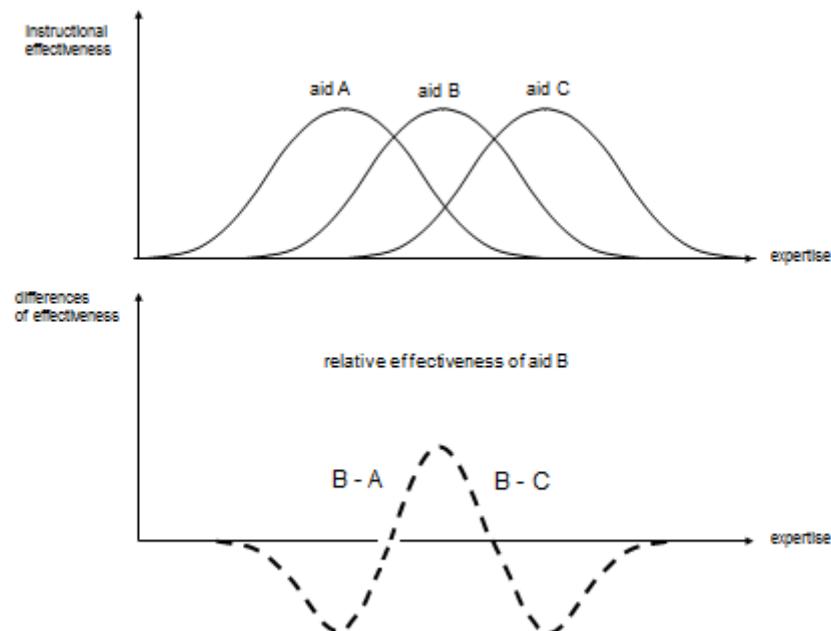
Figure 2. Relative effectiveness of instructional aids depending on the learner's level of expertise.

As for the estimation of a learner's level of expertise, one has to take into account that there are several factors that can partially compensate each other. For example, if visual text is used, lower verbal literacy could be compensated by higher prior knowledge to some extent, whereas this kind of compensation is not needed if auditory text is used. Similarly, visual literacy could compensate verbal literacy to some extent, if pictures and graphs play a major role in the process of learning. This provides the possibility to estimate the learner's level of expertise as a latent variable based on the available manifest variables.

Conditional Use of Multiple Representations

As mentioned above, there is no simple valid rule of thumb for multimedia design or adaptation, such as "Always add pictures to a text," as one might conclude from the multimedia principle. Instead, the use of multiple representations can only be recommended under specific conditions. According to our above considerations, such a conditional rule for combining text and pictures could have the form of a production rule:

(R1)   IF the content is
                 - complex,
         AND
         IF the student has
                 - low prior knowledge,
                 - sufficient visual literacy,

>   - sufficient spatial cognitive skills,
>   - not seen the picture in question yet,
> AND
> IF the picture has
>   - sufficient coherence with the text
> THEN
>   add the picture to the text.

If these conditions are not met, one has to face the possibility of redundancy effects or expertise-reversal effect leading of the opposite of the intended result.

Conditional Use of Modalities

Similarly, there is no valid rule of thumb such as "If you combine text with pictures, use auditory text,'" as one might conclude from the modality principle. Instead, use of auditory text should be recommended only under specific conditions. A conditional rule for combining pictures with auditory text might be the following:

(R2)   IF the picture
>   - is picture is complex or animated,
>   - has not been seen yet by the learner,
> AND
> IF the text
>   - is easy to understand,
>   - includes many semantic connections to the picture,
> OR if the learner
>   - has low verbal literacy,
> THEN
>   use auditory text.

Rule R2 suggests using the advantages of avoiding split attention only if there are sufficient reasons to do so that no detrimental effects are to be expected. A corresponding conditional rule for combining pictures with visual text might be as follows:

(R3)   IF the picture
>   - is picture is simple and static,
>   - has been seen already by the learner,
> AND
> IF the text
>   - is difficult to understand,
>   - includes not many semantic connections to the picture,
> AND if the learner
>   - has high verbal literacy,
> THEN
>   use visual text.

Rule R3 takes into account that when static pictures have a simple visual structure that can be easily held in working memory, split attention plays only a subordinate role and, as a result, the modality principle becomes secondary also. It further takes into account that spoken text is transient, whereas written text is stable and enables better control of processing and, thus, better alignment between perceptual and cognitive processing of verbal information.

Network Activation-Based Multimedia Adaptation

Note that the three rules R1, R2, and R3 presented above address only relatively clear cases where different conditions point toward the same direction. Between these relatively clear-cut conditions, there are many other combinations where the suggestion might be less clear. Let us assume that the considered variables (e.g., prior knowledge, text complexity) are only measured simply in a dichotomous way as either "high" or "low." The conditional statement of the above rules include the following eight variables: prior knowledge, verbal literacy, visual literacy, spatial cognitive skills, complexity of text, picture novelty, complexity of picture, and picture-text coherence. Rule 1 includes six conditions, whereas rules 2 and 3 include five conditions each. Even if these variables are sufficient for creating adaptive multimedia environments, one needs $2^6 = 64$ rules for the decision of whether or not to add a picture, and one needs $2^5 = 32$ rules for the decision of whether text should be presented in the auditory or in the visual modality. If the variables were measured at a three-level scale (high, medium, low), one would need $3^6 = 729$ rules for the decision of whether or not to add a picture and $3^5 = 243$ rules for the decision of whether text should be presented in the auditory or visual modality. A five-level measure of these variables would result in $6^5 = 15.625$ rules for adding a picture and $5^5 = 3125$ rules for choosing the modality. Although storage and processing speed of computers can easily handle high numbers of rules for multimedia adaptation, the specification of these rules by multimedia environment designers could be rather effortful. In the following, we therefore consider an alternative way of creating adaptive multimedia environments based on a kind of connectionist instructional design model operating on the principle of activating or inhibiting specific network nodes according to the conditions at hand.

The interdependencies between the components of a multimedia environment and the characteristics of the learner can be represented by a network where the nodes represent components, features of components, or learner characteristics. Figure 3 shows a simplified cut-out of such a network representing the two components "text" and "picture" as well as the features "auditory" and "visual" for the text as well as the features "static," "animated," and "perceptual signaling" for the picture. Most of the components and features are combinable. This is, of course, true for the two components "text" and "picture," but also for "perceptual signaling" and "static" or "animated" because signaling can be applied to both static and animated pictures. Even the features "static" and "animated" are not necessarily mutually exclusive because there is a continuum between all parts of a picture being static to most or all parts of the picture being animated. However, the features "auditory" and "visual" for the text are considered mutually exclusive because one should not present the same text simultaneously in the visual and the auditory modality (Mayer, 2009). Instead, one should decide for one or the other modality with respect to a specific text segment.
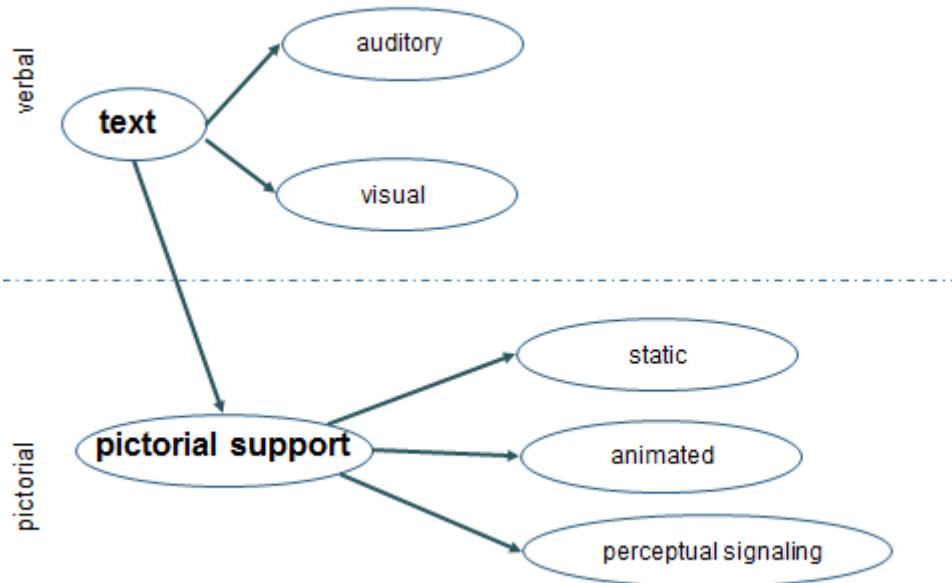
Figure 3. Sample from a dependency network of a multimedia environment. Nodes represent multimedia components or features of these components.

Further unfolding a network of multimedia components and their features means that the network shows a higher granularity. That is, additional nodes are needed to represent conditions that specify the appropriateness of the components or features. Figure 4 shows a snapshot of such a fine-grained constrained activation network around the component representing pictorial support. The nodes around the pictorial support node represent complexity of content, prior knowledge, visual literacy, spatial skills, text-picture coherence, and picture novelty. Their activation can be increased or decreased depending on the degree of the corresponding variable. The connections between these surrounding nodes and the pictorial-support node are either excitatory (symbolized by solid arrows ending with "➔") or inhibitory (symbolized by dashed arrows ending with "➖●"). If a connection is excitatory, an activated emitter node (e.g., "complexity of content") will increase the activation of the receiver node (e.g., 'pictorial-support'). If it is inhibitory, an activated emitter node will dampen the activation of the receiver node. Connections can have different weights according to the strength of the influence of the emitter node on the receiver node.
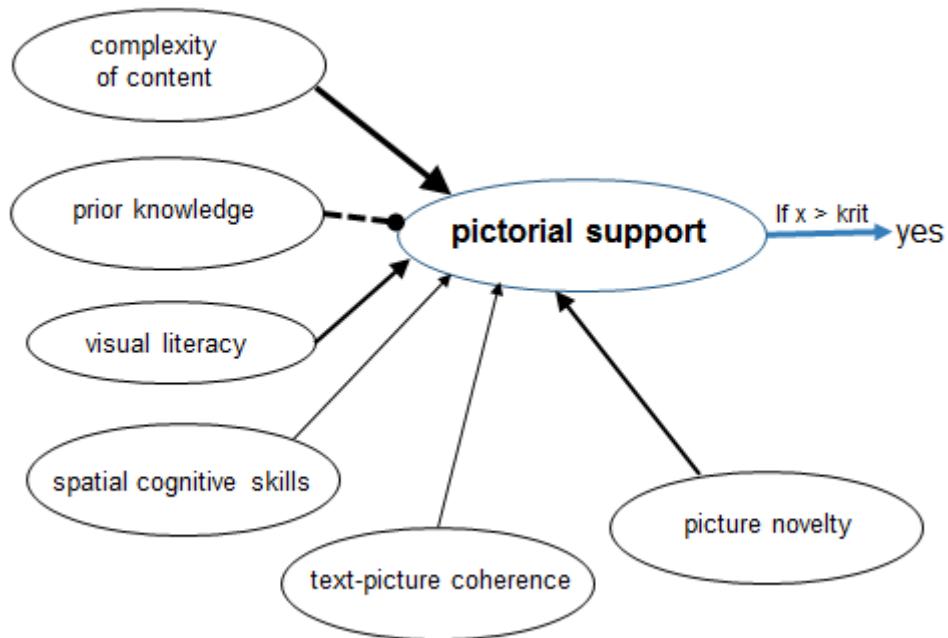
Figure 4. Constrained activation network for determining pictorial support in a multimedia environment.

There is only one inhibitory connection in Figure 4, namely, between prior knowledge and pictorial support, which means that the higher the learner's prior knowledge, the lower the need for pictorial support. All other connections are excitatory: the higher the complexity of the content, the higher the learner's visual literacy and spatial cognitive skills, the greater the semantic connects (coherence) between text and picture, and the higher the novelty of the picture for the learner is, the greater the reasons for pictorial support (i.e., adding a picture to the text). Technically speaking, one can assume that if the activity of the pictorial-support node exceeds some threshold, this indicates the need for including a picture in the present case.

Another example of a fine-grained constrained activation network is shown in Figure 5. This network deals with the decision of using auditory text or visual text when combined with a picture. If picture complexity and picture novelty for the learner is high, if there are many semantic relations (coherence) between the text and the picture, if text difficulty is low and if the learner's verbal literacy is low, activation of the auditory text node will increase. This mirrors the fact that under these conditions, split attention between text and picture will become an obstacle for learning, and therefore, supports auditory text. If the opposite is true (i.e., low picture complexity, low picture novelty, low text-picture coherence, but high text difficulty and high verbal literacy of the learner), activation of the visual text node will increase. This mirrors the fact that under these conditions, split attention between text and picture is not an important issue, whereas cognitive control of text processing is important due to the high text difficulty, which is a critical factor for visual text. Because auditory text should not be duplicated by visual text and vice versa, the two modalities are mutually exclusive with regard to the same text segment. Accordingly, the activity of the auditory text node inhibits the activity of the visual text node, and vice versa.
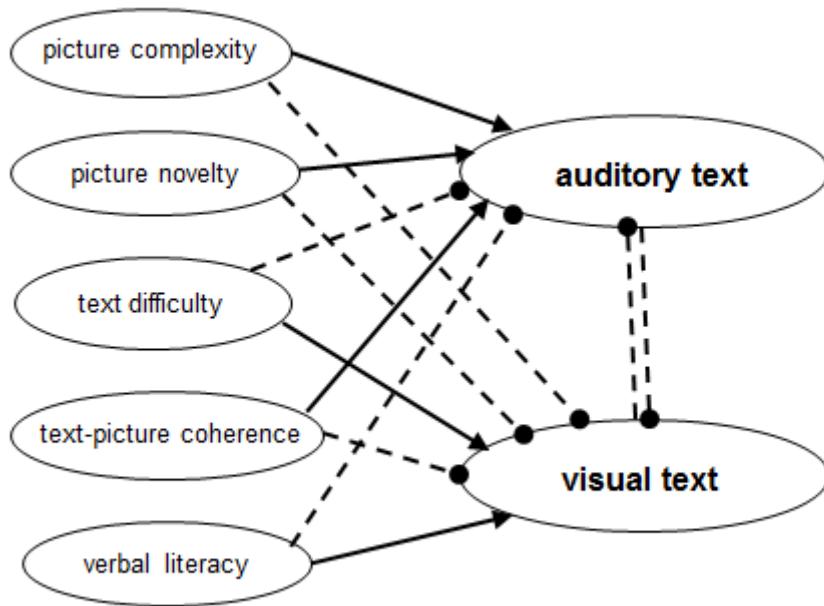
Figure 5. Constrained activation network for determining text modality in a multimedia environment.

The argument could be made that the connectionist constrained activation network for multimedia design and adaptation could better cope with the close to infinite number of conditions that have to be taken into account by an adaptive multimedia environment than a huge set of production rules within a rule-based approach. Whether one or the other principle is more useful for adaptive multimedia environment in terms of implementation and maintenance will be seen in the future.

Recommendations and Future Research

The combination of verbal and pictorial information is a key issue in instructional design and especially in ITSs. Guidelines for the different facets of such combinations will probably be a central part of the pedagogical module within the GIFT system. Up to now, research on multimedia learning has only produced lists of principles (as, for example, the multimedia principle and the modality principle). However, it is not sufficient to derive checklists of simplistic rules of thumb from these principles as such rules apply (if at all) only under specific conditions. Current suggestions for multimedia design can be summarized as follows (Mayer, 2009, in press; Schnotz, in press):

- Use text combined with content-related pictures, when learners have low prior knowledge, but sufficient cognitive abilities to process both the text and the pictures.

- Use pictures only when they are clearly semantically related to the content of the text.

- If written text is used, present it in close spatial proximity to the picture.

- If spoken text is used, present it in close temporal proximity to the picture.

- Do not combine text and pictures if learners have sufficient prior knowledge and cognitive ability to construct a mental model from one source of information, as an additional source would be redundant for them.

- When pictures are used and learning time is not limited, split attention becomes less important. In this case, one should balance the advantage of auditory text (i.e., avoidance of split of attention), which predicts a positive modality effect, against the possible advantage of written text (i.e., higher control of cognitive processing), which predicts a reversed modality effect.

- If the text is difficult to understand, learning time is not limited, and picture complexity is low, use written text rather than spoken text.

- Do not add written text that duplicates spoken text combined with pictures.

- Do not present a text that is semantically related to a picture before the picture can be observed by the learner.

- If the subject matter can be visualized by different pictures in different ways that are informationally equivalent, use a picture with the form of visualization that is most appropriate for solving future tasks.

Up to now, research has only tentatively considered the interdependencies between the corresponding instructional strategies. Because the application of one strategy sets constraints on the applicability of other strategies, the strategies have to "communicate" within tutorial system as conceptualized in the GIFT framework.

We have considered two ways of implementing instructional strategies into adaptive multimedia environments – a rule-based approach and a connectionist constrained network activation approach. Both approaches are, for the time being, only very general ideas that have to be further elaborated and implemented on a pilot basis to assess their functionality, strengths, and weaknesses. In both cases, there are further problems that have to be solved such as, for example, the scaling problem. For some variables such as verbal literacy, spatial cognitive skills, or text difficulty, there are reasonable assessment tools available (such as reading assessment scale, cognitive skills tests, or CohMetrix). For visual literacy or text-picture coherence, the development of instruments is still at the beginning. There are challenges with regard to the complexity of content and prior knowledge, because these variables are interdependent (e.g., Sweller, van Merrienboer & Paas, 1998) and highly domain-specific. Some variables such as picture novelty will probably need to be measured based on simple ratings, such as subjective cognitive load scales and the previous experience of learners. It may also be possible to design a self-adaptive tutorial system, which learns from its instructional practice based on learners' feedback or learning outcomes. Such a system could learn the best practices by reinforcing successful rules and weakening less successful ones or by strengthening some connections in a connectionist network while weakening others. The interplay between research on multimedia learning, development of corresponding instructional strategies, and the assessment of the feasibility and effectiveness of such strategies might become one of the most stimulating fields within the learning sciences during the next decade. GIFT promises to be an excellent framework for implementing and testing these challenging and exciting tasks in the context of ITSs.

References

Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education, 33*, 131-152.

Bernard, R. M. (1990). Using extended captions to improve learning from instructional text. *British Journal of Educational Technology, 21*(3)*,* 215-225.

Chandler, P. & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*, 151-170.

Graesser, A. C., McNamara, D. S., Louwerse, M. M. & Cai, Z. (2004). Coh-metrix: analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers, 36*(2), 193–202.

Hannus, M. & Hyönä, J. (1999). Utilization of illustrations during learning of science: textbook passages among low- and high-ability children. *Contemporary Educational Psychology, 24,* 95-123.

Hildyard, A. & Olson, D. R. (1978). Memory and inference in the comprehension of oral and written discourse. *Discourse Processes, 1,* 91-117.

Hochpöchler, U., Schnotz, W., Rasch, T., Ullrich, M., Horz, H., McElvany, N., Schroeder, S. & Baumert, J. (2013). Dynamics of mental model construction from text and graphics. *European Journal of Psychology of Education, 28*(4), 1105-1126.

Kalyuga, S., Chandler, P. & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 126-136.

Kirsch, I. & P. B. Mosenthal, P.B. (1990), Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*(1), 5-30.

Kulhavy, R. W., Lee, B. J. & Caterino, L. C. (1985). Conjoint retention of maps and related discourse. *Contemporary Educational Psychology, 10*, 28-37.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32,* 1-19.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge, NY: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia Learning* (2nd Ed.). New York: Cambridge University Press.

Mayer, R. E. & Moreno, R. (1998). Split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320.

Mayer, R. E. & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual coding theory of multimedia learning. *Journal of Educational Psychology*, *86*(3), 389-401.

McDonald, T. & Thornley, C. (2002) Unlocking meaning in text: Some thoughts on the literacy challenges our students face. *English in Aotearoa, 48*, 54-60.

McNamara, D. S., Graesser, A. C., McCarthy, P. M. & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, M.A.: Cambridge University Press.

Mokros, J. R. & Tinker, R. F. (1987). The impact of microcomputer-based labs on children's ability to interpret graphs. *Journal of Research in Science Teaching, 24*, 369-383.

Moreno, R. & Mayer, R. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358-368.

Peterson, D. (1996). *Forms of representation*. Exeter, U.K.: Intellect.

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), *Artificial intelligence and the future of testing* (pp. 73-126). Hillsdale, NJ: Erlbaum.

Rubin, D. L., Hafer, T. & Arata, K. (2000). Reading and listening to oral-based versus literate-based discourse. *Communication Education, 49*(2), 121-133.

Schnotz, W. & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction, 13*, 141-156.

Schnotz, W. (in press). Integrated model of text and picture comprehension. In R.E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (2nd Edition). Cambridge: Cambridge University Press.

Schnotz, W. & Bannert, M. (1999). Einflüsse der Visualisierungsform auf die Konstruktion mentaler Modelle beim Bild- und Textverstehen [Effects of the visualization form on the construction of mental models in picture and text comprehension]. *Zeitschrift für experimentelle Psychologie, 46*, 216-235.

Sweller, J., van Merrienboer, J. & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10,* 251-296.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Weidenmann, B. (1989). When good pictures fail: An information-processing approach to the effects of illustrations. In Mandl, H & Levin, J. R. (Eds.), *Knowledge acquisition from text and pictures* (pp. 157-170). Amsterdam: Elsevier.

Wood, D., Bruner, J. & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Child Psychiatry*, *17*, 89-100.

CHAPTER 24 – Support in a Framework for Instructional Technology

Paula J. Durlach
U.S. Army Research Institute for the Behavioral and Social Sciences

Introduction

Digital instructional technology can be designed to adapt education and training to provide each individual learner with a unique experience that is tailored to their aptitude, prior learning, interests, and goals. While there are a number of potential ways in which digital technology could adapt, most systems typically implement only a few. Durlach and Spain (2014) proposed a Framework for Instructional Technology (FIT), which lays out various ways of implementing mastery learning, corrective feedback, and support using digital technology. The framework provides terminology that allows researchers, developers, and designers to characterize instructional systems with greater precision than merely labeling a system as adaptive. FIT can be used to specify precisely how it is adaptive. In FIT, mastery learning is broken down into two separate components, micro-sequencing and macro-sequencing. Micro-sequencing applies to situations in which a given mastery criterion has yet to be met, and a system must determine what learning activity will best promote mastery of the current goal. It can roughly be equated with remediation. Macro-sequencing applies to situations in which a mastery criterion has just been reached and a system must determine the new mastery goal, and what learning activity to provide next. It can be equated with progression to a new topic or deeper level of understanding. For each of the four system behaviors (micro-sequencing, macro-sequencing, corrective feedback, and support), FIT outlines five different methods of potential implementation. Except for macro-sequencing, the five methods of implementation fall along a continuum of adaptation. At the lowest level (Level 0), there is no adaptation – all students are treated the same. Each successive level is increasingly sophisticated with respect to the information used to trigger a system's adaptive behavior.

This chapter focuses on FIT's Support category. Support in FIT refers to guided instruction that is provided during the course of a learning activity. Support can encompass many different types of content and artifacts, such as diagrams, calculators, attentional cues, hints, pumps, and encouragement. In FIT, Support includes any mechanism intended to address student impasse or to lower cognitive load, without changing the task itself (which is dealt with under micro- and macro-sequencing). While FIT focuses on the triggers of support, it says little about its content. In this chapter, we elaborate on support content and review the empirical evidence for the benefits of one FIT Support level over another. The chapter concludes with a discussion of gaps in empirical findings and recommendations for instructional support in GIFT.

Support in FIT

Support in FIT refers to guided instruction that is provided via technology during the course of a learning activity. If highly adaptive, support can be equated with scaffolding – described by Wood, Bruner, and Ross (1976) – as a process that enables a learner to solve a problem, carry out a task, or achieve a goal which would be beyond his or her unassisted efforts. As learner competency increases, scaffolding should gradually be withdrawn (Pea, 2004; Wood & Wood, 1999). This is often referred to as fading. Just as a person healing from a leg injury may go from crutches to a cane to no assistance, ultimately, the learner should be able to apply their knowledge to solving a problem with progressively less assistance. Conversely, just as forcing an able-bodied person to walk with crutches might impair their walking, providing

more support than a learner requires can actually impair their learning (Kalyuga, 2007). Scaffolding thus depends on knowledge of a student's evolving competence, and by its very nature is adaptive. As outlined by FIT, however, not all forms of support are equally adaptive. Puntambekar and Hubscher (2005) make a distinction between "scaffolds" – artifacts and resources, which are not adaptive – and "scaffolding," involving ongoing diagnosis, tailored support, and fading. FIT makes intermediate distinctions between fixed "scaffolds" and adaptive "scaffolding." The different levels of FIT Support are listed in Table 1.

Table 1. The five levels of Support specified in FIT.

Level 0	No support
Level I	Fixed hints on request (problem determined); other fixed sources of information (e.g., glossary) where student initiates access
Level II	Locally adaptive hints, prompts, or pumps a. on request b. triggered
Level III	Context-aware adaptive hints, prompts or pumps (true scaffolding) a. on request b. triggered
Level IV	Same as Level III, with interactive dialog

Level 0 represents no support. Level I represents support that is supplemental, pre-scripted, and accessed on the student's initiative ("scaffolds" in Puntambekar and Hubscher's terms). Examples are a glossary, a study guide, advance organizers, hyperlinks to additional explanatory information, or a "request a hint" button. In the latter cases, information accessed through the link or button is problem-centric. This is sometimes referred to as context-sensitive help (Aleven et al., 2003), where the context referred to is the problem context (not the context of the learner's past performance). All students given that specific problem (or content page) have the same support available for that problem (or page). This can be contrasted with non-context-sensitive help, such as a glossary, which can be accessed at any time, regardless of where the student is in the curriculum.

Level I Support is adaptive only in the sense that the student has the option to use it or not. Cognitive tutors (Anderson et al., 1995) that constrain step order in step-based problem solving typically implement this form of support by providing a fixed sequence of context-sensitive hints. On each problem step, the student can request multiple hints, with each successive hint providing increasingly more specific direction. The final hint – the bottom-out hint – provides the solution if necessary (Guo, Heffernan & Beck, 2008; Roll et al., 2011).

Level II represents resources that are somewhat more adaptive. The support available depends on some aspect of the student's most recent task performance. The support can be either requested by the student (IIa) or triggered automatically (IIb), or both. FIT refers to this as locally adaptive because it takes into account the recent performance (and inferred competency) of the learner. For example, the physics ITS, Andes, can provide "What'sWrong Help" tailored according to the type of error a student commits on a problem step (VanLehn et al., 2005). Andes's Level IIa three-hint sequence is composed (typically) of an attentional cue first (e.g., check your trigonometry), a teaching hint next (stating an abstract version of the knowledge to be applied in the problem), and finally a bottom-out hint, which tells how to correct the error (e.g., replace cos with sin). Sometimes Andes will ask students questions before responding to a help request, in order to better determine the nature of the help to provide. Andes pops up a menu or dialog box for students to supply answers to such questions. In addition to requesting help to correct errors, students can ask Andes for input about what to do next (Next Step Help). The help provided takes into account the principles of physics already applied by the student to this problem (if any).

The ITS, Quadratic (Wood & Wood, 1999), is also an example of IIa. In Quadratic, the level of direction provided in a requested hint is contingent on the student's performance and the level of hint they received on the previous problem. Success on problem N causes the first requested hint on problem N+1 to be given with less detail than the hint for N, whereas lack of success on problem N causes the first requested hint on problem N+1 to occur at a more detailed level than before.

Level IIb Support is given automatically when a system detects some predetermined event. For example, in AnimalWatch (Arroyo et al., 2000) hints are triggered when a student commits an error. Another example of IIb would be sending an alert or hint when a student neglects to perform a desired action. For example, playing the role of a tactical action officer in a scenario-based exercise, a student may be offered a hint to take a particular action if they have not already done so, to avoid imminent danger. Just like for on-request hints, triggered hints may be repeated with increasing direction as time goes by without the student taking the appropriate action (e.g., the first hint might be, "that plane is getting out of radar range," whereas a more directive hint would be "use your drone to keep tabs on that plane").

Inq-ITS is a science inquiry ITS, which employs Level IIb triggered support (Sao Pedro, 2013). Students are given explicit goals to conduct investigations with a simulation by formulating hypotheses, collecting and interpreting data, and drawing conclusions. The triggered support can be followed up by on-request help. For example, if Inq-ITS detects that a student is not collecting data to test their hypothesis, they may receive a text message such as, "It looks like you did great at designing a controlled experiment, but let me remind you to collect data to help you test your hypotheses." The student can respond to the message by choosing an "OK" button or can seek further help by selecting a "How do I do that?" button. Constraint-based reasoning is used to trigger the support, when the system has sufficient data (i.e., after the student has run the simulation at least two times) or when they opt to analyze their data. Violations of a constraint (e.g., never changing simulation variables while running an experiment) trigger a message suggesting a corrective strategy. Multiple violations of the same constraint produce triggered support that is increasingly directive, analogous to increasingly directive hint sequences in cognitive tutors.

Another application that uses Level IIb triggered support is BiLat (Kim et al. 2009). BiLat was designed to provide practice in bilateral negotiation. A student is assigned a mission (e.g., convince the doctor to move the clinic to a different location), and after a research and preparation phase, the student conducts meetings with a series of simulated characters to achieve the mission. During meetings, the student selects speech acts or actions from a menu in order to interact with the simulated characters who react to the selection. During meetings, unsolicited hints are given regarding what would be an appropriate action (in a text window). These hints are triggered according to a model that takes into account meeting phase (e.g., greeting and rapport building, business, closing), available actions, and learning objectives. The first hint for any particular learning objective is given at an abstract level (e.g., "it would be good to begin with a sign of respect"), but is later given at a more directive level if the student does not take an appropriate action ("you should take off your sunglasses"). The coach also provides feedback after a student action (positive or negative).

Level III Support is responsive to an even deeper understanding of student competency than Level II. It is true scaffolding, in which the student is given a tailored level of support to overcome an impasse. To calibrate the level of support appropriately, it is necessary to have an estimation of the student's knowledge and competency. FIT refers to this as context-aware because Level III takes into account knowledge of the learner's performance over time – more than just the current and/or last problem. Rather, Level III Support uses knowledge about the student accumulated over multiple exercises and/or sessions. Suppose a student tactical action officer has already demonstrated mastery of tactical use of a drone. Then triggered reminders for that tactic might be turned off (in which case, omissions might be noted in after-exercise feedback only). Another example would be providing different on-demand hints to a sailor learning to navigate by magnetic compass for the first time versus as refresher training. The

former may be guided through the steps of applying local variation and deviation to calculate the required compass heading, whereas the more experienced seaman may simply be given a standard mnemonic (e.g., TV Makes Dull Children).

EcoLab (Luckin & du Boulay, 1999) is an application that used Level IIIa Support in one of its implementations. EcoLab was intended to support learning about the food web, providing students with the ability to construct a microworld of plants and animals. The system prompted students to perform activities to foster concept learning. The activities could vary among different levels of complexity and abstractness, from very simple and concrete (e.g., the relationship between two specific organisms, such as hawk and a mouse) to more complex and abstract (e.g., a web of relations among herbivores, omnivores, carnivores, and plants). In the Level IIIa implementation (called VIS in the paper), the system determined the difficulty level of activities, based on a Bayesian overlay student model. While on-request help was available at five levels of concreteness, the student model determined what level to give. Thus, when a student requested help, the response was tailored for each student, based on their past history of interactions with the system. If the level of help offered was not at the most concrete level, students could subsequently request more help (ask for the next Support message in the sequence). Since the level of support ultimately delivered was not necessarily the same as the level the student model had predicted as needed, the level actually delivered contributed to updating the student model. It also contributed to determining when to move to a different level of activity difficulty. If the student required the most concrete types of support, the difficulty of the next activity would be decreased, whereas if they need little support, it would be increased (this is considered sequencing in FIT).

FIT Level IV Support "inherits" all the adaptive capabilities described for Level III, and in addition provides delivery of adaptive support via mixed initiative natural language. In the ideal, this allows natural and flexible interaction between the student and one or more pedagogical agents about the learning domain, mirroring the patterns of interaction that occur between human tutors and students (Chi et al. 2001; Graesser, Person & Magliano 1995). During these interactions, the instructional system processes student-generated text or speech, adheres to the social conventions and pragmatics of conversation (such as turn-taking), and provides learning support adapted to the current model of the student.

AutoTutor is an instructional system that supports mixed-initiative dialogues on problems requiring reasoning and explanation in subjects such as physics and computer science (D'Mello & Graesser, 2012). For each problem, AutoTutor's conversations follow a five-step pattern: 1) AutoTutor asks the main question (e.g., When you turn on the computer, how is the operating system activated and loaded into RAM?), 2) student gives initial answer, 3) AutoTutor provides feedback, 4) AutoTutor and student interact to improve the student's answer, and 5) AutoTutor verifies that the student understands. In Step 4, AutoTutor uses hints and prompts to help the student fill in the missing concepts. This is accomplished with the use of a problem script, which contains the ideal answer, a set of expectations about what students will say, potential hints and prompts (and the responses they are aimed to elicit), a set of misconceptions and responses to each one, a set of key words (and their synonyms), and a summary. AutoTutor tracks certain aspects of performance across problems. Factors such as verbosity, answer quality, and overall performance on prior problems influence how feedback and hints are provided by the conversational agent on the current problem. AutoTutor does not, however, use a knowledge component-based student model across problems. For example, on a problem applying Newton's third law, AutoTutor's hinting and prompting will be the same, regardless of the student's prior demonstrated ability to apply Newton's third law, *per se*. Likewise for Why2Atlas (VanLehn et al. 2002), an ITS coach that helps students write essays on qualitative physics, and SCoT-DC, a tutor for shipboard damage control (Pon-Barry et al., 2006). The use of a natural language interface should not necessarily be equated with Level IV Support for all aspects of the tutoring interaction.

Chi et al. (2011) designed a natural language tutoring system, which would exemplify Level IV Support, if it were fully automated. Their purpose was to address a research question: whether tutorial dialogue tactics impact student learning. As answering that question might be undermined by the system's inability to perfectly understand student-generated input, Chi et al. (2011) employed a human to interpret student language. The interpreter's sole job was to determine which item, from a list of potential responses, best matched the actual student response. The list item selected was used by the system as the student input. The human interpreters did not make any tutorial decisions. The tutorial decisions were made by a policy model created by applying reinforcement learning to a physics tutorial dialogue corpus previously generated using a similar system (with tutoring support tactics, such as to whether to provide information or elicit information, selected randomly). Changes in mastery resulting from those dialogues were used as the reward function during reinforcement learning. This machine learning was applied to derive two separate models, one intended to maximize student learning (NormGain) and one intended to minimize student learning (InvNormGain). The two models were then empirically tested with new students, to determine their actual effect on student learning (of introductory college physics). The models were applied in the context of an ITS, where all other factors, such as feedback, were held constant, except for the tutoring support tactics. Both conditions improved student performance from pre-test to post-test; however, the NormGain model produced significantly greater improvement, with effect sizes greater than 0.5 on seven of the eight knowledge components covered. In post-hoc comparison, the NormGain model also produced greater learning than the random model used to generate the original corpus.

The derived NormGain model prescribed tactics that were knowledge component-specific. For example, all other factors being equal, the model might prescribe the tactic of telling information to the student if discussing one knowledge component, but of asking information of the student if discussing a different knowledge component. The derived model included tactics that took into account contextual information such as normative problem difficulty and episodic tutor-student interactions (e.g., how verbose the tutor had been so far) to determine whether to use a tell or elicit Support tactic. A subset of the policies also included conditions concerning student performance. These conditions were sometimes global (e.g., involved performance measures computed from all previous student-system interactions and knowledge components), and sometimes session-specific and knowledge component-specific (e.g., involved performance measures only from the current session and the current knowledge component). The resulting collection of policies for the NormGaim model suggests that student performance is just one of several features that should influence support tactics. Besides taking into account student performance, factors such as prior interactions (e.g., verbosity) and content-specific information (i.e., problem difficulty) are important considerations in providing support during natural language interactions. Moreover, not all knowledge components included a student performance measure in the associated policy (Chi, 2013).

Support Triggers vs. Support Content

The FIT Support levels focus on the information used to trigger support, but they are silent about the nature of the support provided. Examining human tutoring, Chi et al. (2001) identified many forms of tutor-initiated intervention including content-free prompting ("what next?"), hinting (providing a mnemonic), asking leading questions ("when the variation is west we …?"), highlighting critical features (e.g., is the variation east or west?), decomposing a task into parts or steps ("We need to first calculate the magnetic course, then the compass course."), initiating steps ("To get the magnetic course, you need the true course and the variation."), executing steps (10 degrees – 90 degrees = 280 degrees), referring to examples, and real-time correction. Some of these are not deemed as support in FIT (e.g., real-time correction would be deemed corrective feedback, and referring to examples would be deemed micro-sequencing). Nevertheless, the list highlights the variation that is possible in support content.

Chi et al. (2011) focused on the distinction between support content types "Tell" vs. "Elicit." Tell provides the student information, whereas Elicit requests information from the student. They also discussed a specific form of Elicit – Justify, which is the special case of requesting the student to justify or explain their latest step. Various ITS systems have experimented with ways of asking students to justify problem-solving steps (Aleven & Koedinger, 2002; Aleven et al. 04; VanLehn et al., 2002) because such explanation is believed to improve learning (Chi, de Leeuw, Chiu,& LaVancher, 1994; Conati & VanLehn, 2000; Aleven et al., 2004).

Segedy, Loretz, and Biswas (2013) offered a refinement of "Tell" by distinguishing Suggestions and Assertions. Suggestions steer student behavior, whereas Assertions communicate information. The distinction between a Suggestion and an Assertion may not always be entirely clear; however, these might be better thought of as two ends of a continuum of directness. The most direct, Suggestions, tell the student exactly what to do (e.g., "add the deviation to magnetic"). The least direct, Assertions, simply provide information and leave it up to the student to infer how to use that information (e.g., "When the variation is west, it gets added to true."). In between, suggestions may steer without providing the required solution, per se, and may be posed in the form of a leading question (e.g., "when the variation is west we …?"). This example illustrates that the distinction between Tell and Elicit is not entirely clear cut. A "leading question" is leading because it both conveys information (e.g., that east or west matters) and elicits information (what is done when it is west?).

Another distinction highlighted by Segedy, Loretz, and Biswas (2013) was between cognitive and metacognitive support. Cognitive support is aimed at building understanding in the learning domain, *per se*, whereas metacognitive support provides support that is domain general, such as strategy selection and skills for self-regulation of learning (including goal setting, planning, self-monitoring, self-assessment, and reflection). The notion that students can benefit from both cognitive and metacognitive support has been made frequently (Aleven et al., 2006; Azevedo et al., 2011; Dabbagh & Kitsantas, 2012; Roll et al., 2011; Roscoe et al., 2013; Sánchez-Alonso & Vovides, 2007; Shapiro, 2008; Wagster et al., 2007; Zimmerman & Labuhn, 2012). Students with high metacognitive abilities tend to be more effective learners (Chi et al., 1989; Chi et al., 1994; Johnson & Mayer, 2010), but many students are negligent in performing self-regulatory skills (Aleven et al., 2003; Clarebout & Elen, 2009; Shute & Gluck, 1996; Winne & Nesbit, 2009). In the context of providing support in computer-mediated learning environments, researchers have found that students do not necessarily make good use of on-demand help. They may either neglect to ask for help when it could be beneficial, or overuse or "abuse" it to get to the bottom-out hint (Aleven et al., 2003). The observed variation in students' use of on-demand context-sensitive hints led to the development of the Help Tutor (Roll et al., 2011). The Help Tutor, integrated into a cognitive tutor, provides students with feedback on help-seeking errors (such as rapidly requesting multiple hint levels to reach the bottom-out hint, without reading the intermediate hints). The model underlying the Help Tutor recognizes what would be effective or ineffective help seeking-behavior based on the student knowledge component model built up in a cognitive tutor environment. When a help-seeking error is committed, the Help Tutor presents feedback aimed at correcting the error (e.g., "No need to hurry so much. Take your time and read the hint carefully. Consider trying to solve this step without another hint.").

Besides designing support to assist students in overcoming an impasse or improve self-regulated learning, it has been suggested that support should also be designed to bolster student motivation, self-efficacy, and self-esteem (del Soldato & du Boulay, 1995; D'Mello et al., 2010). Several methods have been used to gauge student affect, attitude, and confidence, including questionnaires, analysis of help-seeking behavior, response latency, verbal cues (in natural language systems), eye tracking, facial expression, and body posture. Affective AutoTutor is a version of AutoTutor that uses sensors to detect student boredom, confusion, and frustration, and react to these detected states according to a set of production rules (D'Mello et al., 2010). Both the quality of student input and inferred affective state are used to influence

the content of tutor utterances and the facial expression of the tutor agent. Initial evaluation of the effectiveness of Affective AutoTutor versus the standard AutoTutor in the domain of computer literacy produced complex interactions. Therefore, further work is required to fine tune the policies for providing affective support (D'Mello & Graesser, 2012). The reader interested in learning more about the role of affect and motivation should refer to Section I of this volume.

In the remainder of this chapter, we examine the empirical evidence regarding the learning impact of support in digital instructional technology. There is fairly good evidence that including some form of support in instructional technology is beneficial for learners (Aleven & Koedinger, 2002; Aleven et al., 2003; Clarebout & Elen, 2009; Kali & Linn, 2008); however, few experiments have attempted to compare the effect of support provided in different ways, such as the different levels described by FIT. Our aim is to review the current state of the evidence and determine what is known regarding whether higher levels of adaptivity in support provide sufficiently robust improvements in learning to justify the additional effort and resources needed for their implementation. The review will confine itself primarily to cognitive support. Section I of this volume goes into detail on affect and motivation, while Section II goes into detail on metacognition and SRL.

Empirical Evidence

The first question is whether there is evidence that any kind of support is better than none (Level 0 vs. higher levels). While this seems like a relatively simple question, the answer is ambiguous. Students vary in their tendency to use support provided, and research indicates that whether some support is better than none depends on how the support is used (Aleven et al., 2003; Clarebout & Elen, 2009; Shute & Gluck, 1996). When access to support is under the control of the learner (e.g., via clicking a link or requesting a hint), the first requirement for use is that the learner be aware of support availability. A second requirement is that they be aware of their need for support (or not). This may require self-regulatory skills that vary across learners (Aleven et al., 2003; Clarebout & Elen, 2009; see also Section II on self-regulated learning in this volume). A third factor is learner motivation. Interactions among goal orientation (e.g., desire to gain competence vs. complete a required task), perceived penalties for using support (will it affect the score?), and the perceived value of the support given (is it actually helpful and relevant?) likely affect the learner's tendency to use available on-request support. Aleven et al. (2003) reviewed the empirical literature and concluded that effective use of on-demand help is correlated with better learning outcomes, but additional studies were needed to establish a causal relation. There are some experiments that have shown a causal relation with fixed sources of support. Renkl (2000) found that students provided with on-demand help for solving probability problems performed better in a post-test, compared with students with no help access during learning. In another experiment (Clareboout & Elen, 2009), students learning about obesity performed better on a post-test when they had on-request help (dictionary, instructional goals, example questions, and help interpreting figures and text), compared to other students who did not have that support available. Advance organizers (introductory materials intended to provide structure and summarize concepts to be learned) have also been shown to facilitate learning (Gurlitt et al., 2012; Stone, 1983). A final example comes from students learning science inquiry methods in the context of phase change with Inq-ITs. Students who received blended triggered and on-request support for data collection activities later exhibited better data collection skills in a subsequent no-support condition on a new science topic (free fall), compared with students who had not received support during phase change (Sao Pedro et al., 2013).

Some recent work has used educational data mining to investigate the impact of support; however, the conclusions drawn about whether support helps learning seem to depend on the specifics of the data modeling approach used (Beck et al., 2008; Duong et al., 2013; Goldin, Keodinger & Aleven, 2012; Goldin, Keodinger & Aleven, 2013; Sao Pedro, 2013). Taking into account student use of support does

seem to improve the predictiveness of Bayesian knowledge tracing models (Sao Pedro, 2013; Duong et al., 2013); however, interpretation of some of this work is preliminary, as it focuses only on within-problem use of support or short-term effects. There are multiple instances where learning interventions can boost short-term performance, but actually undermine long-term performance, (Bjork & Bjork, 2011; Dutke & Reimer, 2000; Nückles, Hübner & Renkl, 2008; van Merrienboer, De Crook & Jelsma, 1997). Consequently, inferring lasting benefits from short-term improvement is somewhat problematic.

Relatively few experiments have investigated the effects of using the knowledge of the learner to adapt cognitive support (i.e., few experiments have compared Levels I, II, and III using instructional technology). Luckin and du Boulay (1999) conducted an experiment comparing three implementations of EcoLab. These corresponded to FIT Levels I, IIa, and IIIa (named NIS, WIS, and VIS, respectively, in Luckin & du Boulay, 1999). Twenty-six 10- and 11-year olds completed two sessions with one of the three versions. They completed a written pre-test in an initial session, and an identical post-test in a subsequent session. Level IIIa, described above, presented on-request help with five possible levels of help, with the level presented upon first request determined by the student model. Level IIa presented on-request help, with the level of help determined by the level presented on the previous activity (analogous to Quadratic, Wood & Wood, 1999). Level I presented fixed on-request help, not determined by student past performance. There was a significantly greater learning gain in the Level IIIa condition, compared to Level IIa or Level I. Table 2 lists the effect sizes estimated from the data provided in Luckin and Boulay (1999). Unfortunately, there were other confounded condition differences, which make interpretation problematic. Besides varying how support was provided, the various conditions also differed in how activity difficulty level was determined. In the Level IIIa condition, this was entirely determined by the system (as previously described), whereas in Level I, it was entirely determined by the student. In the Level IIa condition, it was determined by the student; however, the system made suggestions as to the most appropriate level. Because of these confounded differences, it is not possible to determine if the group differences were due to the personalization of support or the method of selecting activity levels (or both). Likewise, Kao & Lehman (1997) also attempted to compare different methods of adapting support, in the context of learning statistical hypothesis testing; however, their experimental conditions had other confounded factors, such as whether feedback was provided on a step-wise or problem-wise basis, and whether students had the opportunity to correct errors mid-problem. Consequently the better post-test outcomes observed for their most adaptive Support condition cannot unambiguously be attributed to the manipulation of support. Similar issues exists with experiments by Azevedo and colleagues (Azevedo et al., 2004; Azevedo et al., 2005).

**Table 2. Comparison of effect sizes in the three conditions
in the Luckin and du Boulay (1999) experiment.**

Conditions	Effect Size (Hedge's g*)
I < IIa	0.37
I < IIIa	1.33
IIa < IIIa	0.60

Some experiments have examined the interaction between individual differences and the content or format of support. The purpose of these experiments has been to examine whether different students should be given different types of support based on relatively stable factors like gender or stage of cognitive development. This is often referred to as macro-adaptation (Park & Lee, 2004). Several analyses on this topic have been conducted with different versions of an arithmetic ITS, AnimalWatch, and to a lesser degree with the geometry tutor, Wayang Outpost (Arroyo, Woof & Beal, 2006; Arroyo et al., 2001 Arroyo et al., 2000, Arroyo et al., 2003). In these studies, students were randomly assigned to different support conditions, and then the data were analyzed to determine if student performance or

attitude was affected by the type of support received, depending on factors such as their gender, stage of cognitive development, spatial ability, and other measures of math skills. Support types varied in factors such as concreteness and interactivity, depending on the study. Results of these analyses were complex, with interactions among gender, other individual difference factors, and support form. Arroyo, Woolf, and Beal (2006) concluded that students at a concrete stage of cognitive development should be first taught with support expressed in concrete terms, but that it is desirable to progress students to more formal and abstract representations as they transition to higher stages of cognitive development. They suggested that it is premature to advocate specific support strategies based on gender alone. Conceivably, providing all learners with multiple forms of support might be a better approach than trying to predetermine which type of support best serves different categories of learners. On the other hand, this presupposes learners will sample the multiple forms and gravitate toward those most effective for them. This is not necessarily the case. Jackson and Graesser (2007) found that students preferred progress feedback over content feedback; however, students learned more with content feedback.

Discussion and Recommendations for Future Research

It is generally agreed that instructional support should be tuned to the current competency of the student – only as much as needed to get over an impasse. Providing too much support (e.g., information already well known by the student) is thought to be detrimental, because it increases cognitive load, leads students to be less attentive, and/or reduces opportunities for new learning. Given that students vary in their effective use of on-demand support and don't necessarily choose options that maximize their learning, why is on-demand support more common than triggered support? Aleven et al. (2003) suggested that despite potentially poor help use on the part of students, in comparing the ability of a student versus a computer program to know when help is needed, the student probably knows best. On-demand support removes the need for a system to infer when to provide help. Triggered support may be intrusive, mis-timed, or off the mark in terms of the support required. On the other hand, the students who need help the most, might be the ones least able to make good decisions about how to use on-demand support. It would seem that a blended approach, combining both on-demand and triggered support might be a feasible solution. Combining student models of student knowledge and student models of help seeking may offer a way to support both knowledge acquisition and self-regulated learning at the same time. Rather than providing students with explicit feedback on their help-seeking behavior (as in the Help Tutor), patterns of students' use of support could be used to actually change the way that support is available – i.e., whether it is triggered, on request, or blended. For example, students who seem to do fine with on-request hint sequences could be left to it, whereas help-abusers might be transitioned to only error-triggered support, and help under-users might be transitioned to blended on-request and triggered support. More research is required to determine the benefits of blending on-request and triggered support.

With respect to comparing the effect on learning of different methods of adapting support (FIT Levels I–IV), there is a disappointing dearth of evidence. Despite the overwhelming belief that support should be adaptive, few technology-based instructional systems have attempted to implement anything above Level II Support and few have attempted to compare the effect of different Levels (possibly none in an unconfounded way). With respect to metacognitive support, Koedinger et al. (2009), stated that their results failed to provide evidence favoring adaptive compared to fixed support. It would be beneficial if more attention were given to determining the relative effectiveness of the various FIT Support Levels.

The inclusion of natural language interfaces in ITS has presumably been inspired by the desire to emulate one-on-one human tutoring, along with the type of Support provided by human tutors. Expert human tutors have wide array of instructional tactics, and the ability to switch among them in real time if required. Yet, VanLehn (2011) suggested that it may not be natural language interaction, per se, that makes tutoring (human or artificial) effective, but rather it may be the granularity at which students

receive corrective feedback. Chi et al. (2011) suggested that among human tutors, natural language tutoring systems, and step-based tutoring systems, none are particularly good at making micro-step support decisions. Moreover, their findings (that the best machine learned policies were knowledge component and context-specific) suggest that general policies for providing support (which exist in most ITSs) might not be the best approach.

What are the implications of this review for GIFT? In general, some support is better than none. The extent to which that support should be tailored on the basis of student performance has not been established, however. Therefore, as a testbed for research, GIFT should be designed so that researchers can easily examine the different FIT Levels for their impact on learning outcomes. Second, the issue of whether support should be on-request, automatically triggered, or a blend of the two is not clear either. Therefore, as a testbed for research, GIFT should have the capability to allow researchers to support all three options and explore various rules for blending on-request and automatically triggered Level II and III Support. Finally, the findings of Chi et al. (2011) need some reflection. GIFT aims to institute general support policies in a pedagogical model that is domain-independent, but Chi et al.'s (2011) work suggests that different policies for different knowledge components, even within a domain, might be more effective. GIFT should enable new research, to further support the development of machine-learned policies. It should be designed to allow support policies to vary in such a way as to collect the data required to derive machine-learned policies. Perhaps, the best policy might be flexibility. Rather than having a set policy for each different circumstance, GIFT might be designed to vary the support tactic and learn what works best for each student at any particular time. Just as a great teacher does, GIFT should be able to recognize when a specific tactic isn't working, and try something else.

References

Aleven, V. & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*, 147-179.

Aleven, V., McLaren, B., Roll, I. & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence and Education*, *16*, 101-128.

Aleven, V., Ogan, A., Popescu, O., Torrey, C. & Koedinger, K. R. (2004). Evaluating the effectiveness of a tutorial dialogue system for self-explanation. In J. C. Lester, R. M. Vicari & F. Paraguacu (Eds.), *Intelligent Tutoring Systems, Lecture Notes in Computer Science 3220* (pp. 443-454). Berlin: Springer-Verlag.

Aleven, V., Stahl, E., Schworm, S., Fischer, F. & Wallace, R. M. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, *73*, 277-320.

Anderson, J. R., Corbett, A. T., Koedinger, K. R. & Pelletier, R. (1995). Cognitive Tutors: Lessons learned. *Journal of the Learning Sciences, 4,* 167-207.

Arroyo, I., Woolf, B., Beal, C.R. (2006). Addressing Cognitive Differences and Gender During Problem Solving. *International Journal of Technology, Instruction, Cognition and Learning*, *4*, 31-63.

Arroyo, I., Beck, J. E., Woolf, B. P., Beal, C. R., Schultz, K. (2000). Macroadapting Animalwatch to gender and cognitive differences with respect to hint interactivity and symbolism. In G. Gauthier, C. Frasson & K. VanLehn (Eds.), *Proceedings of the Fifth International Conference on Intelligent Tutoring Systems*. pp. 574-583. Berlin: Springer-Verlag.

Arroyo, I., Beck, J. E., Beal, C. R., Wing, R.E. &. Woolf. B. P. (2001, May). Analyzing students' response to help provision in an elementary mathematics Intelligent Tutoring System. Paper presented at the meeting of the Help Provision and Help Seeking in Interactive Learning Environments Workshop at the Tenth International Conference on Artificial Intelligence in Education. San Antonio, TX.

Arroyo, I., Murray, T., Woolf, B. P., Beal, C. R. (2003) Further results on gender and cognitive differences in help effectiveness. In U. Hoppe, F. Verdejo & J. Kay (Eds.), *Proceedings of the 11th International Conference on Artificial Intelligence in Education.* Amsterdam, ND: IOS press.

Azevedo, R., Cromley, J. G., Moos, D.C., Greene, J A. & Winters, F. I. (2011). Adaptive content and process scaffolding: A key to facilitating students' self-regulated learning with hypermedia. *Psychological Test and Assessment Modeling*, *53*, 106-140.

Azevedo, R., Cromley, J.G. & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, *29*, 344-370.

Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C. & Greene, J. A. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science*, *33*, 381-412.

Beck, J.E., Chang, K., Mostow, J., and Corbett, A. (2008). Does Help Help? Introducing the Bayesian Evaluation and Assessment Methodology. In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (pp.383-394) Berlin, GR: Springer.

Bjork, E. & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. Gernsbacher, R. W. Pew, L. M. Hough, J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York, US: Worth Publishers.

Chi, M. Personal communication, January 6, 2013.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P. & Glaser, R. (1989). Self explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145-182.

Chi, M. T., H., de Leeuw, N., Chui, M. H. & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.

Chi, M., Siler, S., Jeong, H., Yamauchi, T. & Hausmann, R. (2001). Learning from human tutoring. *Cognitive Science*, *25*, 471-533.

Chi, M., VanLehn, K., Litman, D. & Jordan, P. (2011). An evaluation of pedagogical tutorial tactics for a natural language tutoring system: a reinforcement learning approach. *International Journal of Artificial Intelligence in Education*, *2*, 83-113.

Clarebout, G. & Elen, J. (2009). Benefits of inserting support devices in electronic learning environments. *Computers in Human Behavior*, *25*, 804-810.

Conati, C. & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, *11*, 398-415.

Dabbagh, N. & Kitsantas, A. (2012). Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *Internet and Higher Education*, *15*, 3-8.

del Soldato, T. & du Boulay, B. (1995). Implementation of motivational tactics in tutoring systems. *International Journal of Artificial Intelligence in Education*, *6*, 337-378.

D'Mello, S. K. & Graesser, A. C. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems*, *2*, 23:39.

D'Mello, S. K., Lehman, B., Sullins, J., Daigle, R., Combs, R. Vogt, K., Perkins, L. & Graesser, A. C. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. *Intelligent Tutor Systems, Lecture Notes in Computer Science*, *6094*, 245-254.

Duong, H., Zhu, L., Wang, Y. & Heffernan, N. (2013). A prediction model that uses the sequence of attempts and hints to better predict knowledge: "Better to attempt the problem first, rather than ask for a hint." In S. K. D'Mello, R. A. Clavo & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 316-317). Retrieved November 22, 2013, from http://www.educationaldatamining.org/IEDMS/EDM2013.

Durlach, P. J. & Spain, R. D. (2014). Forms of adaptation in a framework for instructional technology (ARI Technical Report 1335). Fort Belvoir, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Dutke, S. & Reimer, T. (2000). Evaluation of two types of online help information for application software: Operative and function-oriented help. *Journal of Computer-Assisted Learning, 16,* 307-315.

Goldin, I., Koedinger, K. & Aleven, V. (2012). Learner Differences in Hint Processing. In K. Yacef, O. Zaïane, H. Hershkovitz, M. Yudelson & J. Stamper (Eds.), *Proceedings of 5th International Conference on Educational Data Mining* (pp. 73-80). Retrieved October 25, 2013, from http://pact.cs.cmu.edu/pubs/goldin-2012-different-learners%20(1).pdf

Goldin, I., Koedinger, K. & Aleven, V. (2013). Hints: You can't have just one. In S. K. D'Mello, R. A. Clavo & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 232-235) . Retrieved 22, November 2013, from http://www.educationaldatamining.org/IEDMS/EDM2013.

Guo , Y., Heffernan, N. T. & Beck, J E.(2008). Trying to Reduce Bottom-out hinting: Will telling student how many hints they have left help? In B. Woolf, E., Aimeur, N. Nkambour & S. Loajoie (Eds.), *Proceeding of the 9th International Conference on Intelligent Tutoring Systems* (pp. 774-778). Berlin: Springer-Verlag.

Graesser, A., Person, N. & Magliano, J. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, *9*, 495-522.

Gurlitt, J., Dummel, S., Schuster, S. & Nückles, M. (2012) Differently structured advance organizers lead to different initial schemata and learning outcomes. *Instructional Science*, *40*, 351-369.

Jackson, G. T. & Graesser, A.C. (2007). Content matters: An investigation of feedback categories within an ITS. In R. Luckin, K. Koedinger & J. J. Green (Eds.), *Proceedings of the 13th International Conference on Artificial Intelligence in Education (*pp. 127–134). Amsterdam: IOS Press.

Johnson, C. I. & Mayer, R. E. (2010). Applying the self-explanation principle to multimedia learning in a computer-based game-like environment. *Computers in Human Behavior*, *26*, 1246-1252.

Kali, Y. & Linn, M. C. (2008). Technology-enabled support strategies for inquiry learning. In J. M. Spector, M. D. Merrill, J. Van Merrienboer & M. P. Driscoll (Eds.). *Handbook of Research on Educational Communications and Technology (3rd Edition)* (pp. 145-161). New York, NY: Lawrence Erlbaum Associates, Taylor & Francis Group.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509-539.

Kalyuga, S., Ayres, P., Chandler, P. & Sweller, J. (2003). Expertise reversal effect. Educational Psychologist, 38, 23-31.

Kao, M. & Lehman, J. (1997). Scaffolding in a computer-constructivist environment for teaching statistics to college learners. Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Kim, J. M., Hill, R. W., Durlach, P. J., Lane, H. C., Forbell, E., Core, M., Marsella,S., Pynadath, D. and Hart, J. (2009). BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. *International Journal of Artificial Intelligence in Education, 19,* 289-308.

Koedinger, K. R., Aleven, V., Roll, I. & Baker, R. S. J. d. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 383-412). New York: Routledge.

Luckin, R. & du Boulay, B. (1999). Ecolab: the development and evaluation of a Vygotskian design framework. *International Journal of Artificial Intelligence in Education, 10*, 198-220.

Nückles, M., Hübner, S. & Renkl, A. (2008). Short-term versus long-term effects of cognitive and metacognitive prompts in writing-to-learn. In G. Kanselaar, V. Jonker, P. A. Kirschner & F. J. Prins (Eds.), *Proceedings of the 8th International Conference of the Learning Sciences 2008.* Utrecht, NL: ICL

Park, O. & Lee, J. (2004). Adaptive instructional systems. In D.H. Jonassen (ed.), *Handbook of Research on Educational Communications and Technology 2nd edition* (pp. 651-684). Mahwah, NJ: Lawrence Erlbaum.

Pea, R. D. (2004). The Social and Technological Dimensions of Scaffolding and Related Theoretical Concepts for Learning, Education, and Human Activity. *Journal of the Learning Sciences. 13*, 423-451.

Pon-Barry, H., Schultz, K., Bratt, E. O.,Clark, B. & Peters, S. (2006) Responding to Student uncertainty in spoken tutorial dialogue systems. In *International Journal of Artificial Intelligence in Education,* 16, 171-194.

Puntambekar, S. & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist, 40*(1), 1-12.

Renkl, A. (2000). Worked-out examples: instructional explanations support learning by self-explanations. *Research Report 139*, Institute of Psychology, University of Freiburg, Germany. Retrieved 11/08/2013 from https://www.psychologie.uni-freiburg.de/forschung/fobe-files/139.pdf

Roll, I. , Aleven, V., McLaren, B.M. & Koedinger, K. R. (2011). Improving students' help-seeking skills using meta-cognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21,* 267-280.

Roscoe, R. D., Segedy, J., Sulcer, B. Jeong, H. & Biswas, G. (2013). Shallow strategy development in a teachable agent environment designed to support self-regulated learning. *Computers & Education*, *62*, 286-297.

Sánchez-Alonso, S. & Vovides (2007). Integration of metacognitive skills in the design of learning objects. *Computers in Human Behavior*, *23*, 2585-2595.

Sao Pedro, M. (2013). *Real-time assessment, prediction, and scaffolding of middle school students' data collection skills within physical science simulations.* (Doctoral Dissertation, Worcester Polytechnic Institute). Retrieved November 25, 2013 from http://users.wpi.edu/~mikesp/pubs.html.

Segedy, J.R., Loretz, K.M. & Biswas, G. (2013). Suggest-Assert-Modify: A taxonomy of adaptive scaffolds in computer-based learning environments. *Workshop on Scaffolding in Open-Ended Learning Environments , 16th International Conference on Artificial Intelligence in Education*. Memphis, TN, USA.

Shapiro, A. M. (2008). Hypermedia design as learner scaffolding. *Educational Technology Research and Development, 56*, 29-44.

Shute, V. J. & Gluck, K. A. (1996). Individual differences in patterns of spontaneous

online tool use. *Journal of the Learning Sciences, 5,* 329-355.

Stone, C. L. (1983). A meta-analysis of advanced organizer studies. *Journal of Experimental Education, 51* (4), 194-199.

van Merrienboer, J. J. G., De Crook, M. B. M. & Jelsma, O. (1997). The transfer paradox: Effects of contextual interference on retention and transfer performance of a

complex cognitive skill. *Perceptual and Motor Skills, 87,* 784-786.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist, 46*, 197-221.

VanLehn, K., Jordan, P., Rosé, C. P., Bhembe, D., Bottner, M., Gaydos, A., Makatchev, M., Pappuswamy, U., Ringenberf, M., Roque, A., Siler, S. & Srivastava, R. (2002). The architecture of Why2-Atlas:A coach for qualitative physics essay writing. In S. A. Cerri, G. Gouarderes & F. Paraguacu (Eds.), Intelligent Tutoring Systems: 6th International Conference (pp. 158-167). Berlin: Springer.

VanLehn, K., Lynch, C., Schulz, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A. & Wintersgill, M. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Aritificial Intelligence in Education, 15*, 147-204.

Wagster, J., Tan, J., Wu, Y., Biswas, G. & Schwartz, D. L. (2007). Do learning-by-teaching environments with metacognitive support help students develop better learning behaviors? In D. S. McNamara & J. G. Trafton (Eds.), *Proceeding of the 29th Meeting of the Cognitive Science Society* (pp. 695-700). Nashville: Cognitive Science Society.

Winne, P. H. & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, .J. Dunlosky & A. C. Graesser (Eds),. *Handbook of Metacognition in Education* (pp. 260 -277). Mahwah, NJ: Erlbaum.

Wood, D., Bruner, J. S. & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100.

Wood, H. & Wood, D. (1999). Help seeking, learning and contingent tutoring. *Computers and Education*, *33*, 153-169.

Zimmerman, B. J. & Labuhn, A. S. (2012). Self-regulation of learning: Process approaches to personal development. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra & J. Sweller (Eds), *APA Educational Psychology Handbook, Vol 1: Theories, Constructs, and Critical Issues*, (pp. 399-425). Washington, DC, US: American Psychological Association.

CHAPTER 25 – The DENDROGRAM Model of Instruction: On Instructional Strategies and Their Implementation in DeepTutor

Vasile Rus[1], Mark Conley[2], Arthur Graesser [3]
[1] Department of Computer Science; [2] Instructional and Curriculum; [3] Department of Psychology
Institute for Intelligent Systems, University of Memphis

Introduction

We present in this chapter an attempt to conceptualize the notions of instructional strategies and tactics in a way that unifies various views held by different research communities. We then use the proposed conceptualization to describe the implementation of strategies and tactics in DeepTutor (Rus, D'Mello, Hu & Graesser, 2013a; Rus, Niraula, Lintean, Banjade, Stefanescu & Baggett, 2013b), the first ITS based on the framework of Learning Progressions (LPs) (Duschl, Maeng & Sezen, 2011). LPs have been proposed by the science education research community as a way forward in science. The proposed conceptualization of strategies and tactics is intended to help guide the implementation of strategies in GIFT (Sottilare, Brawner, Goldberg & Holden, 2012).

The gist of the proposed conceptualization is a dendrogram model of instruction in which strategies and tactics are defined at various level of instruction granularity resulting in a multi-level, hierarchical structure called a dendrogram. The proposed conceptualization also fits the general (dictionary) definition of strategy as "a plan of action or policy designed to achieve a major or overall aim" or "pattern in a stream of decisions" (Henry Mintzberg; from Wikipeda entry on Strategy), and tactics as local decisions about "actions" (to be precise, as local decisions to choose actions) in the plan shaped by the strategy. Strategies can be viewed as affecting the overall shape of the plan of action while tactics are local decisions resulting in actions. The types of tactics, their mixture, and their enactment are conditioned by the global strategy (which could be many) as well as other factors such as learning environment, target domain, and student background.

Probably, the closest definition of instructional strategy as defined in our dendrogram model is Dick and Carey's (1996) definition of an instructional strategy as the process of sequencing and organizing content, specifying learning activities, and deciding how to deliver the content and activities. Our dendrogram model assumes that this process can be carried out at different levels. Example levels include sequencing and organizing content across grade levels (e.g., K–12 level), across courses within a grade (grade level), within a course (course level), inside a lesson (lesson level), or within an activity such as problem solving. That is, we can talk about strategies at all these different levels of instruction granularity. In our narrative here, we emphasize cognitive (processing of target content) and social aspects of learning and pay less attention to other important components of learning such as affect and motivation. This is primarily due to space constraints and our intent to build on our experience with developing DeepTutor, which for now addresses primarily cognitive and social aspects of learning.

The chapter relates, within the space constraints, the implementation of various instructional strategies in DeepTutor to the GIFT recommendations (Sottilare et al., 2012). According to GIFT designers, there is a distinction between strategies, which are general (e.g., ask a question, prompt the learner for more information, or review basic concepts), and tactics, which are enactments of strategies in a specific domain (e.g., ask a question on concept B, prompt the learner for more information on concept B, or review basic concepts for building clearing tasks). The relation between GIFT definitions of strategies

and tactics and other conceptualizations of these terms is also elaborated as opportunities arise. A first comment on the relationship between our ideas presented in this chapter and the GIFT view of strategies and tactics is that all our descriptions of strategies and tactics are domain-independent while GIFT views only strategies as domain-independent.

We conclude the chapter with yet another model, an action model of instructional strategies and tactics, which we call the Fourier model. The basic idea of the Fourier model is that actions taken by tutors and which are directly visible to students are the result of mixing the effects of many individual strategies at multiple levels of instruction. This mixing of the final signal from simpler ones is analogous to how a final, complex signal is obtained, in Fourier analysis, from simpler signals, i.e. simpler trigonometric functions. The ideas described here are presented in the context of conversational tutors targeting science learning. Nevertheless, they have wider applicability.

Related Research

There is a large spectrum of understanding of the terms strategies and tactics in the education-at-large literature, including the education, psychology, and education technologies literatures. These understandings vary from Foshay's (1975) early first-cut suggestion that there is only one instructional strategy to a myriad of other conceptualizations that use terms such as instructional methods, instructional mechanisms, instructional heuristics, instructional approaches, instructional strategies, pedagogical strategies, tutoring strategies, instructional tactics, tutoring tactics, and cognitive principles of learning. Furthermore, some of the reviewed works use the terms of strategy and tactic in underspecified, ambiguous, or interchangeable ways while others attempt to make a clear distinction between strategies and tactics albeit at different levels of instruction granularity. It is beyond the scope of this chapter to investigate thoroughly the differences among the various definitions and use of the terminology. Nevertheless, we review several conceptualizations of these terms by several research groups in two different communities and propose a model that is a first attempt to unify these various conceptualizations.

Education Literature

As already mentioned, according to Foshay (1975), there is only one instructional strategy and many tactics. The sole instructional strategy is to induce a situation for learning, which must meet the following four conditions of learning: drive or motive (student must want something), cue or stimulus (student must notice something), response (student has to do something), and reward (student has to get something). One can argue that this sole strategy of "inducing a situation for learning" is equivalent to Vygotsky's Zone of Proximal Development (ZPD; 1962). According to Foshay, tactics must meet these four conditions to fall within the umbrella of the "acceptable strategy" (note the circularity of the two notions of strategies and tactics). Foshay (1975) analyzes three tactics, called approaches, with respect to satisfying the four conditions: teacher-centered classroom, mastery-learning, and project-centered approach. For instance, Foshay indicates that teacher-centered classroom is weak at motivating students because this condition is limited to "pleasing the teacher" as students' main drive instead of targeting more personal and therefore more powerful drivers. This view is not entirely incompatible with others' use of the terms strategies and tactics. It is just that the granularity level at which strategies and tactics are defined is very broad.

According to Rothwell and Kazanas (1998), an instructional strategy is an overall plan that governs content (*What is taught?*) and process (*How will it be taught?*). Importantly, the strategy must be specified before content (instructional materials) is created. This makes sense to us, because, for instance, particular strategies or aspects of strategies in DeepTutor were implemented at authoring time as opposed

to instruction time. That is, the implementation of some strategies in DeepTutor require steps taken during authoring (off-line), as suggested by Rothwell and Kazanas, in addition to during tutoring (online).

Rothwell and Kazanas (1998) categorize strategies at two levels of granularity, namely, macro and micro. A macro-strategy is an overall plan governing a course or module. A micro-strategy is a specific plan governing a unit such as one lesson. Rothwell and Kazanas (1998) define an instructional tactic as an activity that facilitates a strategy. They suggest that strategies can be considered in different ways. Based on the philosophy of learning and instruction, they differentiate between expository and discovery instructional strategies. Guided discovery, drill and practice, and inductive exposition are some examples of instructional strategies defined on the basis of philosophy of learning. Strategies can also be categorized based on events of instruction (Gagné, Briggs & Wager, 1992), i.e., what happens during learning and what type of learning is intended. The emphasis is on the link between such instructional events and learned capabilities such as verbal information, cognitive strategy, intellectual skills, motor skills, and new attitudes. We agree with the idea of a dual view (if not multiple views) of strategies without further elaborating at this moment.

Rothwell and Kazanas also present recommendations on how to select among many instructional strategies based on four factors: learners, learning outcomes, learning and working environment, and constraints of the instructional design process. Rothwell and Kazanas point out that any strategy can be adopted but that not all strategies work similarly well under various conditions. As we argue later, deciding which strategies to trigger at each step of the tutoring process is a very important task. Our Fourier model regards tutor's actions as the result of many strategies that are combined by tutors through a complex process (which is yet to be fully understood) that involves deciding which strategies to use and when and how to combine the effects of the activated strategies in ways that resonate with the learner. Examples of tactics that they provide and which were taken from Jonassen, Grabinger, and Harris (1990), are help learners organize information, use cuing systems, provide examples, vary lesson unit size, sequence instruction in logical order, and sequence instruction in learning prerequisite order. Rothwell and Kazanas' strategies and tactics are apparently domain-independent, which differ from the current conceptualization of strategies and tactics in GIFT.

A clear distinction between strategies and tactics is made by Merrienboer and Kramer (1987). Instructional strategies are general design plans that differ in their control of students' processing loads. On the other hand, instructional tactics are specific design plans describing methods to reach learning outcomes under specific circumstances. They mention three instructional strategies, which they call the Expert approach, the Spiral approach, and the Reading approach. These strategies were proposed in the context of computer programming instruction. Tactics were defined using a goal-circumstance-method format. Circumstances affect the outcome of methods but cannot be manipulated. Methods are manipulations designed to lead to desired outcome(s) in given circumstances. The goals in this format are skill-oriented, i.e., specifying what skills students would master in the given circumstances if the specified method is used. For instance, a tactic could be the following triplet (goal, circumstance, method): **goal**=mastery elementary programming background, **circumstance**=students have disadvantaged home backgrounds and are 8 years old or younger, and **method**=ask many factual questions that students are expected to answer correctly.

Intelligent Tutoring Literature

Graesser and colleagues (2001) enumerate a number of ideal tutoring strategies such as Socratic method, modeling-scaffolding-fading, reciprocal training, anchored situated learning, error diagnosis and remediation, frontier learning, building on prerequisites, and sophisticated motivational techniques. It should be noted that strategies and tactics are used interchangeable by Graesser and colleagues (2001) who end a paragraph that describes the above strategies with "tutors clearly need to be trained how to use the

sophisticated tutoring tactics because they do not routinely emerge in typical tutoring sessions with untrained tutors." That is, there seems to be an implied similarity of strategies and tactics.

VanLehn, Jordan, and Litman (2007) make a crisp distinction between tutorial strategies and tactics. Similar to Foshay (1975), they mention only one strategy, which is broadly defined based on the structure of the tutoring activity of solving one problem (their system helps students solve physics problems). The strategy consists of two phases: 1) collaborative problem-solving and 2) reflection, which consists of the student and tutor discussing the solution. An alternative to this strategy is the following: prompting the student for a short-essay answer to a problem followed by feedback and showing a worked-out solution (with no dialogue interaction). These two strategies are defined based on the structure (type and sequence of major phases) of the "solving one problem" activity in a tutoring session. One can image a mixture of these strategies too. For instance, as part of implementing a variable instruction strategy one could alternate between solving problems using the collaborative-reflection strategy and using the "prompt for short-essay answer followed by feedback and worked-out solution" strategy. That is, some strategies serve other higher-level strategies. Our dendrogram model tries to capture this hierarchical relationship among strategies. For this example, the dendrogram model suggest that variable practice is a strategy applied at lesson level (which in this case is problem-solving) while the collaborative-reflection and the "prompt for short-essay answer followed by feedback and worked-out solution" strategies apply at the next level of instruction granularity, which would be the solution level in this case.

Tactics, according to VanLehn and colleagues, are micro-level decisions that control brief episodes of tutoring, such as a single step. Note again the granularity as a major differentiation factor. Examples of tactics are tell-or-elicit a step during problem solving, ask for justification of steps, generate feedback, and figuring out the type of question to ask. A policy is a set of actions, e.g., decisions with respect to tactics, that are to be taken by the tutor during tutor-student interactions. Policies that are successful at inducing student learning gains are sought. Importantly, in a more recent paper, VanLehn and colleagues redefine tactics as policies (Chi, VanLehn, Litman & Jordan, 2011). Nevertheless, we adhere to their earlier view of making a distinction between tactics and policies. As we mention later, we propose to use strategies, policies, and tactics to best describe the behavior of their systems and ours. Policies offer a characterization of a sequence of tactics similar to strategies. We define policies as strategies with a bias (as explained later). Policies may correspond to the control layer discussed in Collins, Brown, and Newman (1987).

We do recognize the above framework described by VanLehn, Jordan and Litman (2007), with some alterations, in DeepTutor. For instance, our general strategy has two phases: collaborative problem solving and summary of the solution to the problem, which can be regarded as a worked-out solution. It should be noted that DeepTutor targets for now the domain of conceptual physics at high-school and college level. The second phase, solution summary, is just a summary without any interactive discussion. While the current design in DeepTutor resembles the two-phase strategy in VanLehn and colleagues' design, we do intend to alter (à la VanLehn, Jordan & Litman, 2007) this strategy as we add in new strategies such as variable practice. In fact, instead of simply replacing this strategy, we plan to add new strategies and a strategy selection layer that will dynamically decide which strategy to use and when.

The four tactics mentioned in VanLehn, Jordan, and Litman (2007) are present in DeepTutor as well in various incarnations. The tell-or-elicit tactic, which is supposedly founded on the "scaffolding-modeling-fading" theory (which Collins, Brown & Newman [1987] call a teaching method, by the way), is present in DeepTutor in the following form: "elicit a step if not articulated by the student" and tell. That is, students are expected to mention all the steps of a coherent solution to a physics problem and if one of the steps was not articulated by the student DeepTutor always elicits it. DeepTutor always tells the step after either the student articulated it or the system elicited it. This is to assure that all the steps are in the common ground of the two conversational partners: tutor and student. Similarly, students are expected to justify their responses using physics principles, and if not, the system prompts them for a justification. In

other words, in DeepTutor there are well-defined policies that emphasize students' articulation of the solution and its justification. We will elaborate shortly on what a policy is and its relation to strategies and tactics. For now, we simply note that "self-explanation" is a strategy, which may be implemented through two tactics: tell-or-elicit and ask-for-explanation-or-not (both are binary tactical decisions). When the outcome of a tactic is set a priori (by the researcher or developer) then it becomes a policy, e.g., always ask for an explanation instead of dynamically decide to ask or simply tell. That is, the bias in the enacted strategy generated from (pre-) setting the outcome of certain tactical decisions leads to a policy, e.g., always ask for an explanation. The policy can also be dynamically learned in which case the bias would often by hard to interpret; this is the case when the policies are learned from big data.

We recognize there is a small price to pay for our "elicit a step if not articulated by the student" policy. After all, high-knowledge students may be annoyed by the fact that they have to articulate all steps and justify them. We believe the payoff is worth this risk in our first implementation of DeepTutor. We do intend to explore different policies in the future. To some extent, we already experimented with a new policy that mitigates this risk, specifically by skipping some of the steps that we are confident many students already master. For instance, solutions to particular problems aligned with higher levels of understanding in the Learning Progression used in Deep Tutor have specific steps being optional (students will get credit if articulated but the system will not elicit these steps). The way we implemented this policy required effort at authoring time because solutions to problems aligned with a higher level of understanding in the LP needed be proactively authored in ways to support the implementation of this policy at tutoring time. There is one strong reason for which we made this choice of shifting parts of implementing this policy at authoring time: to maintain the logical coherence of the solution and dialogue. Doing it entirely dynamically is a very complex task, which we explicitly avoided to tackle (for now). Indeed, maintaining the coherence of the dialogue and of the solution itself is challenging. If some steps of the solution are dynamically skipped, there is a high risk of ending up with a broken dialogue and incoherent solution with undesirable effects on learning. It should be noted that our current implementation of this tactic (using VanLehn and colleagues' meaning of tactic) does not account, for instance, for other factors such as students' affective state or motivation. We do plan to add strategies, policies, and tactics that address learners' affect and motivation.

Due to space reasons, we do not elaborate on the other VanLehn and colleagues tactics' implementation in DeepTutor. We would like to add that one advantage of DeepTutor over the system described by VanLehn, Jordan, and Litman (2007) is macro-level adaptation, i.e., the selection of instructional tasks based on students' background. Macro-level adaptation implies the need for macro-level strategies and tactics to address issues such as instructional task selection and sequencing (see Rus et al., 2013a). Furthermore, it should be noted that DeepTutor addresses conceptual physics, while the system described by VanLehn, Jordan, and Litman (2007) includes quantitative problem solving accompanied by conceptual explanations. These differences may further explain the different instantiations of the strategies and four tactics in VanLehn and colleagues' paper.

Collins, Brown, and Newman (1987) present three previously studied "pedagogical methods:" reciprocal teaching, procedural facilitation, and teaching problem solving. Their analysis led to a framework for guiding the design of learning environments. The framework itself includes six teaching methods: modeling, coaching, scaffolding, articulation, reflection, and exploration. Interestingly, the teaching method of coaching involves observing students and offering "hints, scaffolding, feedback, modeling." They also mention "heuristic strategies" that are rules of thumb for how to approach a problem. One such rule specifies how to distinguish special cases in solving math problems. It seems to us that their heuristic strategies may be equivalent to tactics in GIFT due to their domain specificity. Problem-solving strategies and the control strategies (how to select among problem-solving strategies) are part of the content dimension of their framework. The framework includes three other dimensions: method (which includes the six teaching methods mentioned above), sequencing, and sociology. An interesting remark that offers

support for our hierarchical dendrogram model is the fact that the control strategies "operate at many different levels", e.g., across domain problem-solving strategies or more domain-specific heuristics and strategies.

In conclusion of this brief literature review, we remark that one major difference among the various works cited above is the level of granularity at which instruction is analyzed, and consequently, the strategies and tactics defined. Indeed, there are strategies at course level (Rothwell & Kazanas, 1998), lesson level (Rothwell & Kazanas, 1998), activity level (VanLehn, Jordan & Litman, 2007), and step level (VanLehn, Jordan & Litman, 2007). We define our dendrogram model based on these (and some other) levels of instruction granularity. This allows us to preserve the many usages of the terms strategies and tactics, as exemplified by the above literature.

DENDROGRAM Model of Instruction

Based on the previous overview of the various conceptualizations of strategies and tactics, we propose a framework to unify them. The framework is based on the observation that various researchers and communities define strategies and tactics at different levels of granularity. Furthermore, because different researchers use different terms such as principle, approach, method, strategy, tactics, and so on, we propose to use just two terms, strategies and tactics, for all levels of granularity. The alternative would be to use a larger set of terms, each attached to a specific level of instruction granularity. For instance, we can talk about instructional methods at course level, about strategies at lesson level, and tactics at solution level. The difference between using the shorter or larger set of terms is not conceptually important as the emphasis in both cases is the same: identifying strategies (or being called principles or methods or else) that shape the types, frequency, and sequence of activities as well as the delivery format (overall called plan of action). The tactics would be the activities or local decisions regarding the activities (which result in actions) included in the resulting plan.

As already mentioned, our proposed framework defines strategies and tactics at different levels of instruction granularity. For instance, following Rothwell and Kazanas (1998), there is a course level of instruction as well as a lesson level. Strategies at course level are about organizing (choose topics, their frequency/repetition-rate, and sequence them) the sessions/lessons in a whole course. Examples of such strategies are strategies based on LPs (as in DeepTutor) or strategies based on topic prerequisite structure. At the lesson level, the strategies indicate how to choose the type and frequency of activities and sequence them within a single lesson, e.g., following a guided inquiry based instructional strategy (another typical strategy at this level is variable practice which entails presenting the target concepts in different contexts). The next level of instruction granularity is activity (within or associated with a lesson). A typical activity within a lesson is problem solving, i.e., applying learned concepts through problem solving, which is the standard activity in DeepTutor (Rus et al., 2013a) or the tutor described by VanLehn, Jordan & Litman (2007). In DeepTutor, there is an outer-loop that manages the problem-solving activity at macro level and which is responsible for choosing the problems and sequencing them appropriately. Once a problem is selected, then DeepTutor selects the steps of the solution (this is called the solution level of instruction granularity). Specific strategies and tactics can be defined at this level (VanLehn, Jordan & Litman, 2007). Furthermore, in DeepTutor, there is a sequence of hints associate with each step in the solution to a problem. DeepTutor uses a relatively complex mechanism to choose the type and best sequence of hints to give students in order to help them articulate the step by themselves. There could be strategies and tactics defined at this step level as well. The solution and step levels of instruction granularity in DeepTutor correspond to VanLehn's (2006) inner-loop although in our case the solution and step levels of instruction are implemented as two nested loops. The inner loop in VanLehn (2006) two-loop framework manages the student-system interaction while the student works on a particular task, e.g., solving a physics problem. For reference, the outer-loop in VanLehn's framework selects the next task for the

student to work on. In summary, based on the above analysis of our own work and others' work we can talk about strategies and tactics at course level, lesson level, activity level, solution level, and step level.

There is one extra level above course, which we call standard/curriculum level of instruction, which is meant to create a plan for covering a target domain across grade levels. Strategies at this level shape the major topics and their sequencing across grade levels. A typical strategy at this level would be based on Learning Progressions that map the instruction of big ideas in a domain across grade levels. Other strategies would be to use the prerequisite structure or recommendations from experts. One could argue for yet another level, the grade level. Strategies at grade level shape the mixture and sequences of courses within a grade. For instance, an Introduction to Physics course should be taught in Fall and an Advanced Physics course in Spring. Figure 1 offers a simple view of the hierarchical levels of instruction, which can be associated with specific strategies and tactics.
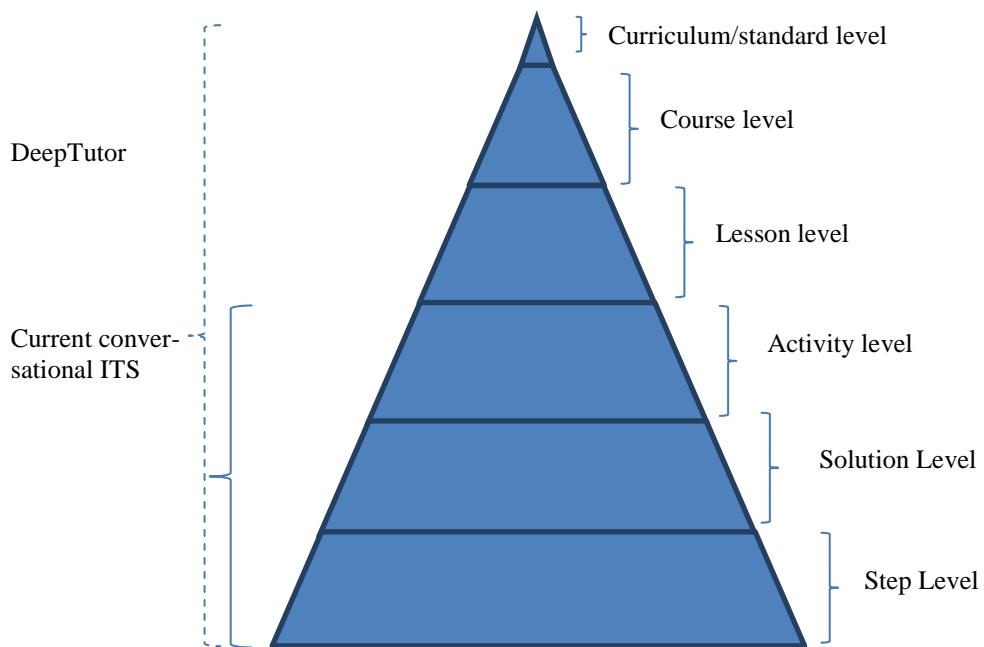
Figure 1. Simplified view of the hierarchical dendrogram model of instruction.

Figure 2 shows a more detailed view of the proposed general model of instruction, which takes the shape of a dendrogram, an hierarchical structure used in data clustering to show the relationship among clusters of various granularities. We show at each level the major activities that are included in the plan at that level, which is, in turn, shaped by the corresponding strategies. For instance, at lesson level, the strategies will shape the lesson plan, which means choosing the type, frequency, and sequencing of activities within a lesson. Note that the lesson plan includes types of activities, frequency, and an actual sequence of activities. The same strategy can result in different mixtures of types (activity 1, activity 2 - Problem Solving, activity 3), frequencies (30%-40%-30% mixture of activity 1, activity 2, and activity 3, respectively), and sequencing. The different mixtures are the result of applying different tactics during actual instruction. The tactics thus make local decisions (e.g., trigger a deep question next or not) based on dynamic aspects of the learning environment and student background and input. The strategy imposes constraints on the type of tactical decisions thus determining the space of possible plans/mixtures. A particular sequencing that is chosen for a particular student or groups of students with same characteristics during one session is a policy. In general, a policy is a strategy with a bias. A policy can be defined by the developer, determined at training time, or learned, e.g., through data mining.
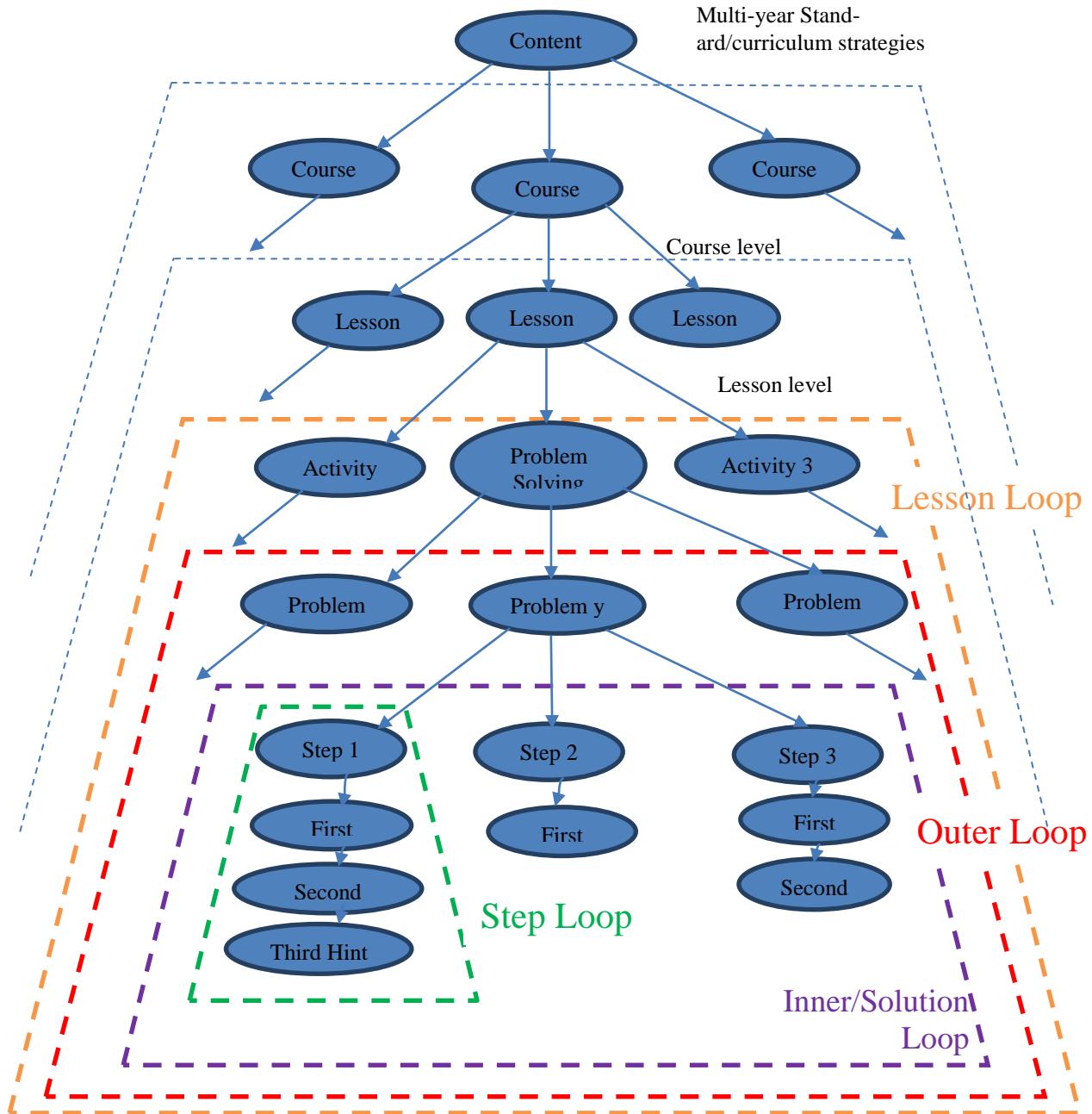
Figure 2. Detailed view of the dendrogram model of instruction illustrating the various levels of instruction granularity, the major activities/actions, and strategies.

The three concepts of strategy, policy, and tactics can be applied to all levels of instruction granularity. For instance, when creating a lesson plan, a teacher may use a guided inquiry instructional strategy to generate the plan, which means specifying the activities, frequency, and their sequence. During actual instruction she may use tactics to adjust the plan based on the dynamics of the classroom resulting in an enacted lesson plan (Gunckel, 2008). There is a variety of possible enacted lesson plans and the guided inquiry strategy delimits the space to a subset. That is, the strategy defines possible plans while during actual instruction through tactics and based on the dynamics of the learning environment, including student responses, an actual plan will be enacted. The actual plan will be known only once the class is over. Sometimes, the bias in the plan is clear and known. For instance, an "self-explanation" strategy

could be biased toward "an always ask for an explanation" policy. That is, the bias in the strategy is generated from (pre-)setting certain tactical decisions, leading to a policy. The policy can be also dynamically obtained, e.g., through reinforcement learning, which often may result in a policy without a clear interpretation of its bias (explained next); this is also the case when the policies are learned from big data.

We analyze now how these three concepts (strategy, policy, and tactic) apply at the activity level based on the design in VanLehn, Jordan, and Litman (2007). According to VanLehn, Jordan, and Litman (2007), they learn a "strategy" (at solution level according to our dendrogram) from data. The strategy is grounded on the modeling-scaffolding-fading "theory." It should be noted that such learned strategies are anonymous as often there is no clear interpretation for each learned sequence of "elicit-or-not" and "explain-or-not" actions. In fact, it would be hard to find a label for each mixture of long sequences of the two tactics as the space of possibly mixtures is large. Using our strategy-policy-tactic concepts, we can regard their design as involving one strategy (modeling-scaffolding-fading) that defines the set of possible mixtures of tactics. A particular mixture of tactics is a policy (learned from data in their case) that may bias the overall plan (all the actual actions taken by the tutor during the solving of one problem) in one way or another. For instance, at one extreme, a resulting policy would be to always elicit, while at the other extreme, it would be to always tell. Yet another policy would be "elicit a step if not articulated by the student" and "tell", i.e., articulate the step to make sure it is in the common ground of the conversation. This is the current policy in DeepTutor at the problem-solving/activity level. The policy is more sophisticated as we have step types. For instance, we have optional steps, which are steps that are not necessary for a good solution to a problem but which, if articulated, need be acknowledged. It should be noted that when eliciting a step from the student, DeepTutor offers appropriate scaffolding, as needed, to help the student articulate the missing step. The scaffolding is done using a constructivist strategy that encourages students to articulate the step by themselves. Help is offered only when the student struggles. Sequences of progressing hints are available for each step in the problem that provide less information initially (more vague hints) and then progressively more information, depending on the particular student. The strategy at this step level is more complex than that. For instance, we have conditional hints, which will only be triggered when certain conditions are met. Other strategies to scaffold the articulation of a step in the solution to the physics problem could be used.

In the dendrogram model in Figure 2, strategies at higher levels impact the strategies and tactics or the implementation of strategies at lower levels. For instance, a guided inquiry-based strategy for generating lesson plans at the lesson level would entail certain types of strategies, policies, and tactics at lower levels and exclude others such as lecturing. Certain activities in a plan entail their own specific strategies. For instance, there is a general problem-solving strategy that can be summarized as follows: 1) read and understand the problem carefully, 2) identify the given and unknown variables, 3) generate a strategy to find the unknown variables based on what is given and also based on world and domain knowledge, and 4) execute the strategy to find the unknown variables in order to solve the problem.

A question arises with respect to what levels in the dendrogram ITS developers should worry about. The answer differs depending on the perspective: current state of the art versus long-term vision. Current ITSs mainly address the problem-solving activity level, solution level, and step level. Early attempts to dynamically address the course/standards and lesson levels were made during the design of DeepTutor. Indeed, because DeepTutor relies on LPs, which track big ideas related to a target domain across grade-levels, we can claim that DeepTutor covers all levels (our LPs are aligned with curriculum standards). Addressing all the instruction levels in our dendrogram model will allow future educational technologies to serve three major types of users: 1) aspirational users whose learning needs are expressed as professional goals ("I would like to become an electrician"), 2) conscious learners who are aware of their need-for-improvement with respect to a topic (their needs are expressed in the form "I would like to learn more about Newtonian physics") and 3) focused learners who have a specific learning need, e.g., need help with solving a concrete problem whose solution is due tomorrow. A one-stop-shop educational portal,

e.g., with a simple interface à la Google where learners simply type their needs, would handle these three types of users differently by directing them to educational technologies that implement strategies at different levels of granularity. For instance, aspirational users would be directed to a computer tutor that implements curriculum and standards-level strategies that will be able to generate a study/degree plan based on student's background. The study plan will include a list of courses in an appropriate sequence. Once the degree plan has been generated, the student will be directed to specific computer tutors that handle the course level of instruction. Current ITSs can handle mostly the conscious learners. Human tutoring services such as Tutor.com can handle the third type of users as current ITSs are not yet ready to help students with "surprise" problems, which are problems unknown a priori to the computer tutor.

Instructional Strategies in DeepTutor

DeepTutor (Rus et al., 2013a; Rus et al., 2013b) is the first ITS based on the framework of LPs (Duschl, Maeng & Sezen, 2011). LPs organize a target domain from a learner's perspective and the basic idea is to map out successful learning trajectories that students follow from naïve conceptions to mastery. The LPs are modeled as a set of hierarchical levels of understanding which increase in sophistication the higher up in the hierarchy they are. We define an instructional trajectory as a sequence of instructional tasks that students are engaged in and which are meant to help students develop sophisticated conceptualizations of the target domain, i.e., move up the LP hierarchy.

We distinguish in DeepTutor (Rus et al. 2013a) among strategies that are appropriate at the macro level, i.e., strategies that help decide the type and sequencing of instructional tasks a student is supposed to work on during a single training session or across many sessions, and at the micro level, which are strategies that impact the interaction between the learner and DeepTutor within a task, e.g., while solving a particular problem. Examples of macro-level strategies are anchored learning and spacing. They correspond to the lesson level in our dendrogram model of instruction. Examples of micro-level strategies are question asking and feedback. These strategies correspond to the solution and step levels, which together correspond to the inner-loop in VanLehn's two-loop framework. DeepTutor includes macro-level strategies corresponding to the activity level and above (e.g., LPs can be viewed to span many sessions, i.e., the course level, many courses (grade level) and grades (curriculum standard levels) in the dendrogram model while the micro-strategies correspond to the solution level and below). As the dendrogram shows, there are strategies at the step level also that control the atomic or sub-step loop, which in DeepTutor consists of sequences of hints that have types and are dynamically sequenced. Generally speaking, there is one constructivist scaffolding strategy at the atomic or sub-step loop: help students articulate the step by providing minimal information, i.e., only as much as students need in order to articulate the answer by themselves. There is a quite sophisticated mechanism to guide the step-level scaffolding, e.g., hints have types and are dynamically sequenced based on the student model while maintaining the coherence of the solution to the problem and dialogue.

We paid special attention to one level of instruction granularity in DeepTutor: course level. Because we have designed and integrated in DeepTutor, an LP for Newtonian physics, DeepTutor is able to sequence the topics in a course on Newtonian physics across many sessions. In other words, there is a course-level loop in DeepTutor. However, the nature of the interaction between the suggested domain map in the LP and the activity of problem solving is not yet fully understood and therefore the advantages or disadvantages of various course-level strategies are not yet fully known. As of this writing, we have experimented with two course-level strategies: 1) a drilling/depth strategy in which over multiple sessions of DeepTutor-student interaction students are drilled on a major topic per session, e.g., Newton's third law, and 2) a breadth strategy, which was used in one-session experiments; the idea was to remediate the weakest aspects of students' knowledge across all Newtonian physics topics in one session. Our Newtoni-

an physics LP (Rus et al., 2013a; Rus et al., 2013c) includes seven strands corresponding to seven major topics in Newtonian physics such as Newton's first law, Newton's second law, free-fall, etc.

Using existing terminology in the literature, we can claim that DeepTutor implements a number of strategies including scaffolding through constructivist dialogue, modeling-scaffolding-fading, multiple representations with explanation and coordination, self-explanation, deep questions, modeling (summarization, asking questions), anchored instruction, variable practice, feedback, assessment, interleaving topics versus drilling on one topic (sometimes the choice between two is constrained by experimental setup as mentioned earlier), worked-out examples, and contrastive cases. A general problem-solving strategy is also implemented in DeepTutor because the major activity is problem solving. Domain-specific tactics are used to implement the strategy while generating specific solutions to specific problems. These tactics are primarily enacted at authoring time by experts who generate the solutions to the problems. DeepTutor simply and dynamically reenacts the problem-solving strategy encoded in the solution at tutoring time.

The bottom line is that the actions that DeepTutor takes and that are directly observed by students are the result of many strategies that are simultaneously active. In general, we can say that the actual moves taken by a tutor (human or computer-based) are the result of many strategies addressing the various aspects of learning (socio-cognitive-affective-motivational) under various constraints from environment such as technological, modality inspired, and pragmatic. When exposed to actions that are the result of many strategies, students must decode the actions that they directly perceive in ways that tutors hope would have the best impact on learning. Therefore, it is tutors' job to best select and mix the strategies in ways that students' decoding is most successful from a learning perspective.

To best model this strategy composition and decoding process, we propose a Fourier transform model inspired from Fourier analysis in mathematics according to which a general function can be decomposed into a sum of basic trigonometric functions. The sum of all the functions (see Figure 3) results in the complex function, which in our tutoring context corresponds to the moves the tutor actually makes and which the student then perceives. That is, a tutor does a Fourier synthesis during tutoring, which consists of summing up the effects of individual strategies. In fact, the tutor may do more than just a sum operation because the process of combining strategies is yet to be understood.
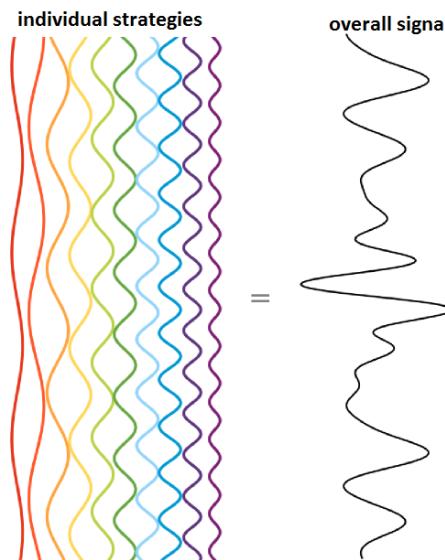
Figure 3. The resulting wave corresponds to a tutor move that the student perceives and is the sum (meaning a complex way of combining) of individual waves that correspond to different strategies and tactics.

As already mentioned, a critical step in our Fourier model of tutor actions is combining the effect of many activated strategies resulting in a specific tutor move, which the student then perceive. We present next two ways to approach this composition of strategies problem: assembly of binary decision processes or a more elaborate process. In the first approach, we assume that at each moment all strategies will be inspected and a decision be made whether to activate or not activate the strategy (a binary decision). Once all the activated strategies are available, a composition step that combines the effect of all the active strategies into one or more actions would follow. The composition could be simple, i.e., a simple concatenation of effects/actions or complex. For instance, the activation of a "coordination of representations" strategy and "induce confusion" strategy could be implemented as a dialogue move through a concatenation of two sentences: one for the "coordination of representations" ("Look at the image.") and a contradictory statement meant to induce a state of confusion ("The large truck exerts a larger force on the small car" when in fact the forces are equal in magnitude – confusion has been shown to help learning [D'Mello, Lehman, Pekrun & Graesser, in press]). Alternatively, the composition could aim at generating more naturalistic dialogue moves that combine the effects of the two strategies into one sentence: "As shown in the image, the large truck exerts a larger force on the small car." While this type of response is more natural and preferable, the underlying composition process is complex. As expected, the complexity increases significantly when more than two strategies must be accounted for.

The other approach, the elaborate process, would be to learn a function that maps the set of strategies into the desired outcome directly. For instance, the learned strategies activation function will simply generate the naturalistic response ("as shown in the image, the large truck exerts a larger force on the small car") in a context in which the right factors would justify such a response (and assuming the function was successfully learned). There is no explicit way to treat separately the two strategies whose effects are embedded in the above response.

We would like to add that the opposite process of observing the actions of a tutor and finding its individual strategies would be equivalent to a Fourier analysis of the synthesized tutor response. The synthesized tutor response (see the overall signal to the right of Figure 3) that embeds/hides the effects of many strategies and is sent over a noisy channel. The tutor response is then transmitted over a noisy channel. That is, what the student actually perceives would be a noisy version of the synthesized response. The role of the instructor is to generate the right signal that best "resonates" with the student. That is, the synthesized signal should be such that it makes it easier for the student to decode it in ways that enable a triggering of effective learning processes.

Recommendations and Future Research

In conclusion, we would like to make several points. First, there is a range of understandings for strategies and tactics in the education-at-large literature. One major difference among the various papers we discussed is the level of granularity at which instruction is analyzed, and consequently, the strategies and tactics defined. We defined a dendrogram model of instruction based on various levels of instruction granularity, which allowed us to account for the many different usages of the terms strategies and tactics. At one extreme, we can claim there is only one strategy in tutoring as observed from the highest level of instruction granularity: maintaining the learner in "the zone" (of proximal development). On the other hand, the actions taken by a tutor can be viewed as the result of many strategies that simultaneously shape these actions. This view is captured in our Fourier analysis model of instruction actions.

Second, based on the terminological disparity among various research groups and communities, there is a need to standardize the use of terminology. The GIFT project is pushing for this unification and our proposal is meant to help in this effort.

Third, many strategies at the same level of granularity seem to overlap or be dependent on each other. An analysis of these commonalities, differences, and interdependencies is needed. For instance, there is an intrinsic interdependency between the strategies of asking questions and guided self-explanations or between asking questions and scaffolding in conversational tutors. Similarly, for instance, Collins, Brown, and Newman (1987) define coaching as being dependent on modeling and scaffolding.

Fourth, some strategies must be re(de)fined. For example, self-explanation is not an instructional strategy but rather a learning strategy (what the learner does). The instructional strategy would be "encouraging, maintaining, and increasing self-explanation behavior." As a general rule of thumb, one could say that any learning strategy can be transformed into an instructional strategy by way of the tutor "encouraging, maintaining, and increasing" student's use of a particular learning strategy.

Fifth, there are other aspects of tutoring that impose their own strategies or constraints on the implementation of core cognitive strategies, such as the ones discussed earlier. Pragmatic aspects impose the use of strategies to, for instance, mitigate gaming-the-system behavior. Modality and technological constraints shape the implementation of strategies. We already mentioned the need to maintain the logical coherence of the solution and of the dialogue in conversational tutoring systems. The role of the instruction delivery method is indeed important (Anderson, 1983). Similarly, technological constraints have an impact on the type of strategies or their implementation. For instance, in dialogue-based ITSs, the assessment of students' utterances rely on NLP techniques. Although much progress has been made in the area of NLP, the current state-of-the-art algorithms are not perfectly accurate. Due to this limitation, in DeepTutor, we have implemented a strategy that avoids one of the worst situations in tutoring, which is giving negative feedback when the student is right. This typically results from a false negative generated from the natural language assessment module because of limitations of the NLP technology. That is, due to these technological limitations sometimes a student response is deemed incorrect even though it is correct. The opposite is also true: sometimes an incorrect response is deemed correct. However, this situation is less harmful because even if the system believes the student is right when in fact the student is not, it provides positive feedback and then asserts the correct answer (which would be different from students' incorrect answer). The bottom line is that in this case the student will eventually see the correct response. The former case is worse because high knowledge students know when they are right and getting negative feedback disengages them, sometimes to the point of losing confidence in the system or, in the worst-case scenario, even quitting using the system. Because we believe the false negatives are worse, we adopted a policy (strategy with a clear bias) that tends to give students the benefit of doubt when the assessment module is less confident. That is, our policy is to give positive feedback when student's response is close to being correct but not sure.

Sixth, strategies that address other aspects of learning, such as affect and motivation, must be considered and their interaction with the major cognitive (targeting processing of content) strategies discussed in the context of GIFT. Similarly, strategies targeting social aspects must be addressed as well. Social strategies, e.g., strategies that modulate the dialogue interaction between the tutor and student, are inherent in our dialogue-based computer tutor DeepTutor. We also have strategies that encourage verbosity or whose goal is perfect grounding at every turn. Furthermore, it could be argued that the providing feedback strategy is by itself touching upon affect and motivational aspects of learning. For instance, a negative feedback in response to an incorrect student statement would clearly impact students' affective and even motivational state.

Another view of strategies, in accordance with the dual (if not multiple) view of strategies similar to Rothwell and Kazanas (1998), would be that strategies are general and that implementing them requires specifying their social, cognitive, affective, and motivational effects. In other words, the implementation of a strategy must indicate its social, cognitive, affective, and motivational impact. Some strategies may have limited or no impact on one or more of these dimensions. One can also argue that the way a strategy

is implemented (through a policy), not its core intent, will impact more or less a dimension of learning. For instance, the form of positive feedback can have more or less of an impact on learner's motivation. Indeed, the following three different forms of implementing positive feedback "Good" vs. "Outstanding answer" versus "Perfect answer that is way above your peers" could have different degrees of motivational impact. We plan to further explore this idea of describing each strategy based on their impact on the four main dimensions of learning.

We would like to end with noting that understanding the role of strategies, policies, and tactics in ITSs is a complex issue that is yet to be fully understood. We hope this chapter will help with making progress toward this goal and, in particular, will help with the conceptualization and implementation of strategies and tactics in GIFT (Sottilare et al., 2012).

Acknowledgment

References

Dick, W. & Carey, L. (1996). *The systematic design of instruction*. 4th ed. New York, NY: Harper Collins

Duschl, R., Maeng, S. & Sezen, A. (2011). Learning progressions and teaching sequences: A review and analysis. *Studies in Science Education*, *47*(2), 123-182.

D'Mello, S., Lehman, B., Pekrun, R. & Graesser, A.C. (in press). Confusion can be beneficial for learning. *Learning and Instruction*.

Foshay, A.W. (1975). Teaching Tactics and Teaching Strategy. *Educational Leadership*, *32*(6), 373-5.

Graesser, A.C., Person, N., Harter, D. & TRG. (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, *12*, 257-279.

Gunckel, K. L. (2008). *Pre-service elementary teachers learning to use curriculum materials to plan and teach science* (PhD Dissertation). Michigan State University, East Lansing, MI.

Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher 30,* 141-166.

Jonassen, D. H., Grabinger, R. S. & Harris, N. D. C. (1990). Analyzing and selecting instructional strategies and tactics. *Performance Improvement Quarterly, 3*(2), 29-47.

Rothwell, W.J. & Kazanas, H.C. (1998). *Mastering the instructional design process: A systematic approach*. San Francisco, CA: Jossey-Bass.

Rus, V., D'Mello, S., Hu, X. & Graesser, A. C. (2013a). Recent advances in conversational intelligent tutoring systems. *AI Magazine, 34*(3), 42-54.

Rus, V., Niraula, N., Lintean, M., Banjade, R., Stefanescu, D. & Baggett, W. (2013b). *Recommendations for the Generalized Intelligent Framework for Tutoring based on the Development of the DeepTutor Tutoring Service.* Paper presented at the User Meeting on Generalized Intelligent Framework for Tutoring (GIFT), The 16th International Conference on Artificial Intelligence in Education (AIED 2013), July 9-13, Memphis, TN.

Rus, V., Baggett, W., Gire, E., Franceschetti, D., Conley, M. & Graesser, A.C. (2013c). Towards Learner Models based on Learning Progressions in DeepTutor, Learner Models. In R. Sottilare, X. Hu, A. Graesser & H.

Holden (Eds.) *Design Recommendations for Intelligent Tutoring Systems: Learner Modeling, Volume I*. Army Research Laboratory.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). *The Generalized Intelligent Framework for Tutoring (GIFT)*. Downloaded from www.gifttutoring.org on November 30, 2012.

VanLehn, K. (*2006*). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*. *16*(3), 227-265.

VanLehn, K., Jordan, P., and Litman, D. (2007). "*Developing pedagogically effective tutorial dialogue tactics: Experiments and a testbed*" Proceedings of SLaTE Workshop on Speech and Language Technology in Education (ISCA Tutorial and Research Workshop).

Van Merrienboer, J. & Krammer, H. (1987). Instructional strategies and tactics for the design of introductory computer programming courses in high school. *Instructional Science*, *16*, 251-285.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*. *16*(3), 227-265.

Vygotsky, L.S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

<h1 style="text-align:center">CHAPTER 26 – Scaffolding Made Visible</h1>

Andrew M. Olney
University of Memphis

Introduction

The notion of *scaffolding* in instructional contexts is pervasive. According to the citation analysis tool *Publish or Perish* (Harzing, 2007), there are over 100,000 citations of scholarly works in this area. Despite the popularity of scaffolding, or perhaps because of it, there is substantial disagreement about what exactly scaffolding means: "the concept of scaffolding has become so broad in its meanings in the field of educational research and the learning sciences that it has become unclear in its significance" (Pea, 2004, p. 423). This state of affairs has led some to assert that, "there is no single right answer to what the word scaffolding means" (Sherin, Reiser & Edelson, 2004, p. 388).

This chapter presents a critical analysis of the concept of scaffolding as it has evolved over time. The analysis differs from existing reviews (Stone, 1998a; van de Pol, Volman & Beishuizen, 2010) in at least two ways. First, rather than assuming that scaffolding is a metaphor that needs to be formalized with a normative framework, we closely analyze source texts to uncover the epistemology of scaffolding. Second, rather than exclusively focusing on the origins of scaffolding, we use later work from the original line of research to augment and clarify its origins.

The analysis begins by examining the current controversies surrounding the common notion of scaffolding as a metaphor for support. After reviewing the history of scaffolding and its theoretical foundations, we explore the structure and function of scaffolding. We examine the implementation of scaffolding in human tutoring and intelligent tutoring systems. Our analysis suggests that key components of expertise have a larger role to play in scaffolding than appears in common practice. In particular, we argue that the most effective scaffolding makes expertise visible.

Scaffolding: Just A Metaphor?

In this section, we describe the status of scaffolding as a metaphor, the controversies surrounding the metaphor, and a shift in meaning of the metaphor in the field. A great deal has been written on the subject of scaffolding as a metaphor (Cazden, 1979; Greenfield, 1984; Brown & Palincsar, 1986; Palincsar, 1998; Stone, 1998a, 1998b; Pea, 2004; Quintana et al., 2004; Sherin et al., 2004; Lajoie, 2005; Puntambekar & Hubscher, 2005; van de Pol et al., 2010; Belland, Walker, Olsen & Leary, 2012). In a widely cited review, Stone (1998a, 1998b) examines what he calls the "metaphor" of scaffolding as discussed by Wood, Bruner, and Ross (1976). In the everyday sense, scaffolding is a temporary structure used by workers who are either building or repairing a building. However, using this everyday sense raises a number of questions when trying to understand Wood et al. (1976)'s notion:

> Discussions of problem solving or skill acquisition are usually premised on the assumption that the learner is alone and unassisted. If the social context is taken into account, it is usually treated as an instance of modelling and imitation. But the intervention of a tutor may involve much more than this. More often than not, it involves a kind of "scaffolding" process that enables a child or novice to solve a problem, carry out a task or achieve a goal which would be beyond his unassisted efforts. This scaffolding consists essentially of the adult "controlling" those elements of the task that are initially beyond the learner's capacity, thus permitting him to concentrate upon and complete only those elements that are within his range of compe-

> tence. The task thus proceeds to a successful conclusion. We assume, however, that the process can potentially achieve much more for the learner than an assisted completion of the task. It may result, eventually, in development of task competence by the learner at a pace that would far outstrip his unassisted efforts. (Wood et al., 1976, p. 90)

It's quite clear from this description that instructional scaffolding is a kind of support. But Wood et al.'s description raises a number of questions if we consider scaffolding as a metaphor, because a metaphor invites us to interpret an analogy. For example, one can ask whether the tutor is the scaffolding itself, or if the scaffolding is supporting the tutor and the student as they work together on the building. The same issues of interpretation apply to the building the scaffolding surrounds. It is unclear whether the building is merely a solution to a particular problem, or whether the building represents competencies for solving an entire class of problems. The basic problem in interpreting scaffolding as a metaphor is that different interpretations implicitly make different theoretical commitments regarding the nature of learning and even the nature of cognition. In the example above, one might argue that the tutor can't be an active participant in the construction of student competence because such competence is internal to the student.

Without a theoretical framework to interpret it from, the metaphor of scaffolding is simply too diffuse to mean anything other than support. For this reason, many researchers explicitly connect the notion of scaffolding to Vygotsky's (1978) *ZPD*, although this connection wasn't explicitly made by Wood et al. (1976). The ZPD is a theory that reconciles the relationship between learning and development, namely, that learning precedes development. The ZPD "is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers," such that "what a child can do with assistance today she will be able to do by herself tomorrow" (Vygotsky, 1978, pp. 86-87). In other words, a child's development may be defined by two levels. The first level is defined by the child's independent problem solving – what they can do on their own. The second level, the ZPD, is defined by what the child can do with the assistance of others. Collaborative learning within the ZPD enables a child's development because it leads to future independent problem solving.

Ostensibly one would expect additional clarity to emerge if the metaphor of scaffolding were interpreted in the context of the ZPD. However, clarity remains elusive. For some, the ZPD can be incorporated quite simply, "Scaffolding can help learners accomplish tasks within their ZPD (Vygotsky, 1978) by providing the assistance learners need to accomplish tasks more complex than they could do alone in a way such that they can still learn from that experience" (Quintana et al., 2004, p. 340). Yet aside from introducing the ZPD into the definition, this version of scaffolding seems no clearer than before. In fact, it raises the further question of whether ZPD and scaffolding are simply two ways of expressing the same idea. For others, the ZPD has many implications that necessitate analyzing scaffolding in terms of essential features like contingency, fading, and a transfer of responsibility (Stone, 1998a; van de Pol et al., 2010). These features flow from the ZPD in the following way: the support from the tutor must be tuned to the student's ZPD (contingency), and because the student will inevitably learn, the tutor must reduce support appropriately (fading) so the student can assume a greater role over time (transfer of responsibility). In many ways, these deeper analyses convert the scaffolding metaphor into a theoretical concept, as is further discussed in the next section.

Puntambekar and Hubscher (2005) provides an insightful review of the evolution of the scaffolding metaphor from the original tutoring context of Wood et al. to the modern classroom context of teacher, peers, and artifacts like educational software. Using their own analysis of the essential features of scaffolding, Puntambekar and Hubscher conclude that much progress has been made in terms of providing support. Support can come from many sources (teacher, peers, or artifacts) and technological support is becoming increasingly sophisticated and diverse. However they also conclude that the emphasis on support has been at the expense of defining features of scaffolding like contingency and fading (see

Table 1). A tool that provides a fixed level of support to all students may provide too little support to some students and too much support to others. Any fixed level of support, by definition, is insensitive to a student's ZPD.

Table 1: Evolution of the Notion of Scaffolding. Reprinted with permission from Puntambekar and Hübscher (2005).

Feature of Scaffolding	Original Notion of Scaffolding	Evolved (Current) Notion of Scaffolding
Shared understanding	Adult or expert establishes shared understanding of common goal and provides motivation	Authentic task often embedded in the environment; provides a shared understanding
Scaffolder	Single, more knowledgeable person provides support to complete the task	Assistance is provided; tools and resources
	Multimodal assistance provided by a single individual	Distributed expertise – Support is not necessarily provided by the more knowledgeable person, but by peers as well
Ongoing diagnosis and calibrated support	Dynamic scaffolding based on an ongoing assessment of the learner (individual)	Passive support – Ongoing diagnosis by peers and or software is not necessarily undertaken
	Adaptive scaffolding – Support is calibrated and sensitive to the changing needs of the learner	Blanket "scaffolding" – Support (especially in tools) is the same for all students
Fading	Eventual fading of scaffolding as students become capable of independent activity	In most cases, support is permanent and unchanging

History and Analysis of Scaffolding

In this section, we review the origins of scaffolding and how it relates to the ZPD. Our central claim is that approaching scaffolding as a metaphor is ill-conceived. Instead we argue that scaffolding should be considered as a *label* for a theoretically well-defined phenomenon. The notion of scaffolding has been attributed to Wood et al. (1976) by a number of scholars (Brown & Palincsar, 1986; Wood & Wood, 1996b; Davis & Miyake, 2004; Pea, 2004; Lajoie, 2005), though some have been more cautious in their attribution (Stone, 1998a). Thus we review the depiction of Wood et al. (1976) and analyze it as being authoritative.

In the original description by Wood et al. (1976), scaffolding is not presented as simply a metaphor. Rather scaffolding is presented as a construct defined by a set of interrelated constructs. While the quote of Wood et al. in the previous section explicitly uses "a kind of scaffolding" to introduce the intuitive justification for the label "scaffolding" for a particular construct, this should be considered no differently than the use of "cancer" (from the Latin "crab," so used because of the crablike appearance of blood vessels around a tumor) to refer to that construct. Indeed, it is clearly foolish to speculate how crabs, by analogy, might help us better understand cancer.

In order to justify these claims, we briefly review the methods and findings of Wood et al. (1976). Their method involved a series of trials consisting of a child (ages 3–5 years), an adult tutor, and a task of assembling wooden blocks into a pyramid (the so-called "Tower of Nottingham" [Jones & Ritter, 1998]). The pyramid consisted of six layers. The top layer was a single block. The other five layers each consisted of four interlocking blocks, with each block in a layer of a different shape (A, B, C, D [Jones & Ritter, 1998]). The difference between layers was that each layer was successively smaller, yielding a pyramid shape. Once constructed, layers themselves locked together via a notch on the bottom and a corresponding knob on the top of the layer below. In other words, the problem consisted of creating layers (with each layer having the same type of solution structure) and combining the layers together. Because all blocks in the five layers had the same size interlocking knobs and holes and were the same thickness, it was possible to incorrectly assemble blocks belonging to different layers.

When a child entered the experiment room, the tutor, Ross, allowed them to play with a jumble of the disconnected blocks for 5 minutes. The tutor would then encourage the child the make a pair. From that point on, the tutor would intervene as little as possible in order to let the child perform the task themselves. The tutor would intervene in two kinds of situations. First, if the child could not or would not produce anything, the tutor would demonstrate or present a partial solution. Second, if the child attempted assembly but made an error, the tutor would either ask the child to compare their assembly to a correct assembly or directly correct the error.

In their analysis of the tutor's responses to the children's problem solving, Wood et al. noticed two shifts. The first shift was the relative amount of help needed by each age group. The 3- and 4-year-olds needed roughly the same total amount of help, but the 5-year-olds only needed half that amount. The second shift was the type of help required, either demonstrating (showing) or verbally intervening (telling). The 3-year-old ratio of show to tell was 3/2, but the 4- and 5-year-old ratio was about 1/2. So while the 3- and 5-year-olds were distinct both in the amount and type of help needed, the 4-year olds shared aspects of both: they needed as much help as 3-year-olds, but the type of help needed was comparable to 5-year-olds. This finding led Wood et al. to further elaborate on the intuitive label of scaffolding:

> It is in this sense that we may speak of a scaffolding function. Well executed scaffolding begins by luring the child into actions that produce recognizable-for-him solutions. Once that is achieved, the tutor can interpret discrepancies to the child. Finally, the tutor stands in a confirmatory role until the tutee is checked out to fly on his own. (Wood et al., 1976, p. 96)

If this were the end of their discussion of scaffolding, then it could be presumed that the term was meant largely in a metaphorical sense. However, Wood et al. proceed to not only indicate that their initial description of scaffolding was intended only as a teaser but also to outline an entire theory of scaffolding functions:

> We may now return to the beginning of the discussion. Several functions of tutoring – "scaffolding functions" – were hinted at in the introduction. We can now elaborate more generally upon their relation to a theory of instruction. What can be said about the function of the tutor as observed in this study? (Wood et al., 1976, p. 98)

Wood et al. define six scaffolding functions based on their study. The first function is recruitment. During recruitment, the tutor draws the child into the task by gaining their attention, stimulating their interest, and fostering a level of commitment to the learning task. Through the second function, reducing degrees of freedom, the tutor reduces task difficulty to the appropriate level for the child's ability level. Direction maintenance, the third function provided by the tutor, keeps the child focused on the current goal and provides affective and motivational support when needed. Marking critical features is the fourth function, by which the tutor draws attention to relevant task features, such as highlighting incorrect solutions. The

fifth function is frustration control. Frustration control by the tutor helps reduce negative affect that would impede successful learning. The sixth function, demonstration, consists of modeling the solution. Modeling can consist of a complete demonstration of the task, or a partial demonstration based on the child's current attempt.

It is worthwhile noting that scaffolding is presented as theoretical definition, or construct (Hurley, 2011). Quite clearly, it no longer refers to a metaphor for building, but has now been redefined in terms of a theory of optimizing learning during tutoring. It should further be noted that none of the scaffolding functions themselves are operationalizations. Rather, each function is its own construct. For example, there are many ways to recruit, many ways to mark critical features, and many ways even to model, depending on the task. Thus, although the study itself was limited to a problem-solving activity with wooden blocks, the scope of the scaffolding functions is much broader.

Given this fuller view of scaffolding, what then is the relationship between scaffolding and the ZPD? Again, this relationship is somewhat complicated by a lack of mention of the ZPD in Wood et al. It is quite difficult to understand precisely why the ZPD wasn't mentioned in Wood et al. when the work of the first two authors is considered. The second author, Bruner wrote the introduction to the 1962 edition of *Thought and Language* (Bruner, 1962), which mentions the ZPD by name (Vygotsky, 1962, p. 103), a decade before introducing scaffolding. In his introduction, Bruner also discusses Vygotsky's work using blocks problems in studies with children. Although these blocks problems typically used nonsense words to create categories for perceptually different blocks, at least one of Vygotsky's block problems has notable parallels: 22 blocks in four categories (Fodor, 1999). Finally, in later work, Bruner himself states that scaffolding was an attempt to better specify how the ZPD would work in practice (Bruner, 1986). The first author, Wood, in previously published work, writes of a "'region of sensitivity to instruction' – a hypothetical measure of the child's current task ability and his 'readiness' for different types of instruction" (Wood & Middleton, 1975, p. 181). The similarity of this description, "region of sensitivity to instruction" to the phrase "Zone of Proximal Development" is uncanny. Also in later work, Wood argues that the ZPD underspecifies both the nature of the guidance from the tutor and the learning that takes place in the student (Wood & Wood, 1996b). Thus, both Bruner and Wood's earlier work suggests that they were familiar with the concept of ZPD when they introduced scaffolding, and in their later work, they both portray scaffolding as an elaboration of the ZPD concept.

The assumption that the ZPD is a subtext to the original work of Wood et al., rather than connection made retrospectively, licenses fruitful interpretations and comparisons. In agreement with both Bruner (1986) and Wood and Wood (1996b), we argue that scaffolding is not identical or redundant to the ZPD, but instead offers some a more explanatory framework to an otherwise more descriptive theory. We characterize the elaboration of ZPD by scaffolding as having both theoretical and mechanistic dimensions. The strongest theoretical elaboration is the implied equation of comprehension and production with the endpoints of the ZPD:

> In the terminology of linguistics, comprehension of the solution must precede production. That is to say, the learner must be able to recognize a solution to a particular class of problems before he is himself able to produce the steps leading to it without assistance. (Wood et al., 1976, p. 90)

The theoretical contribution of this description is quite strong when interpreted in terms of the ZPD. According to the ZPD, what children can do independently is their actual developmental level, here equated with production. However, comprehension sets the ZPD itself, because it puts an upper bound on what the child would be able to do even with assistance. If the child is unable to comprehend a solution or partial solution to a problem, then that problem is outside their ZPD. To some extent this characterization is implicit in Vygotsky's discussion of language development and the ZPD (Vygotsky, 1978), but making

it explicit leads to additional hypotheses and insights. Across the literature, comprehension and production are often asymmetric, and comprehension often seems to precede production (Clark & Hecht, 1983). However, task demands, such as asking children using gaze instead of pointing to respond to stimuli, can make this effect appear larger or smaller (Brandt-Kobele & Hohle, 2010). Clark and Hecht (1983) have argued that an essential part of learning language is the coordination of comprehension and production. They suggest that repairs to speech, whether repairs of pronunciation or wording, indicate that children are monitoring their own production. In their terminology, comprehension sets a standard for production. During production, both comprehension and production processes are active, and comprehension processes are used to monitor and repair production errors. This appears to be the characterization of the ZPD that Wood et al. intend: the ZPD is the gap between comprehension and production, such that the "knowledgeable other" provides monitoring and support in production to "coordinate" production and comprehension. Bruner calls this monitoring and support "vicarious consciousness" (Bruner, 1986, p. 74).

Scaffolding also elaborates the ZPD in terms of processes and mechanisms. Perhaps the most insightful is Wood et al.'s explanation of the tutor's behavior in terms of domain and student models. In the decades since, domain and student models have become regarded as defining features of ITSs that emulate human tutors (Woolf, 2008). They write:

> The effective tutor must have at least two theoretical models to which he must attend. One is a theory of the task or problem and how it may be completed. The other is a theory of the performance characteristics of his tutee. Without both of these, he can neither generate feedback nor devise situations in which his feedback will be more appropriate for this tutee in this task at this point in task mastery. The actual pattern of effective instruction, then, will be both task and tutee dependent, the requirements of the tutorial being generated by the interaction of the tutor's two theories. (Wood et al., 1976, p. 97)

The interaction between these models is implicitly the key driver for all the scaffolding functions they go on to define. This mechanistic explanation is an advance over Vygotsky, whose writings on the possible tutor actions were descriptive at best. Some examples of adult guidance or collaboration given by Vygotsky include asking leading questions, demonstrating, and solving problems in collaboration (Vygotsky, 1978). He does not outline when the adult should intervene or what kinds of guidance or collaboration are more appropriate in a given situation. Moreover, the examples given by Vygotsky, at best, correspond to only two of the six scaffolding functions of Wood et al., highlighting critical features (of which asking leading questions is an instance) and modeling (synonymous with demonstration). It appears the that processes and mechanisms of Wood et al. provide a fuller view of the phenomenon across its duration, ranging from recruitment at the beginning, scaffolding, and then the withdrawal of tutor support (also known as fading, see Collins, Brown, and Holum [1991]).

It is unfortunate that the domain of Wood et al.'s work, small children assembling blocks with a tutor, was so distinct from more formal educational domains. The gap between their domain and formal educational domains raises the further question of whether scaffolding, as they define it, is applicable to more practical educational domains. At first analysis, there seems to be reason to think that the theory of scaffolding is generally applicable. The ZPD was defined for classroom and non-classroom contexts, where the "knowledgeable other" could be either an adult or peers (Vygotsky, 1978). If scaffolding were assumed to be merely an elaboration of the ZPD, then it would stand to reason that scaffolding can be properly situated in these contexts. The problem, perhaps, is that scaffolding per se is defined with very fine-grained models of the student and the domain, and it is not clear whether a human teacher could simultaneously monitor the ZPDs of an entire classroom in order to closely adapt instruction (as noted by Puntambekar and Hubscher [2005] above). From this perspective, it may be the case that scaffolding does introduce a restriction on Vygotsky's original conception of the ZPD in that it requires more careful modeling than the ZPD originally warranted. After all, the activities that Vygotsky mentioned as guid-

ance, such as demonstration, collaborative problem solving, and asking leading questions, don't necessarily require a fine-grained model of the student to implement.

A similar question may be asked for other tutoring domains: to what extent is Wood et al.'s conception of scaffolding applicable to tutoring in subjects like research methods or physics? In the blocks problem, everything is visible except the solution structure. The kinds of manipulations, like picking a block up, rotating it, or pushing it into another block, are essentially givens even for small children. In contrast, more formal educational domains have a good deal of invisible or abstract elements. Algebra, for example, requires knowing things, having familiarity with abstract operations, etc. Thus, the question as to whether Wood et al.'s notion of scaffolding is preserved between domains is a non-trivial one.

Several studies of naturalistic human tutoring have revealed the so-called "5-step tutoring frame" (Graesser, 1993; Graesser, Person & Magliano, 1995). These same structures have featured prominently in the work of other researchers conducting fine-grained analyses of tutoring (Chi, 1996; Chi, Siler, Jeong, Yamauchi & Hausmann, 2001). The 5-step tutoring frame begins with the introduction of a problem. As indicated by the name, the following five steps are enacted in order:

1. TUTOR asks a difficult question or presents a problem.

2. STUDENT gives an initial answer.

3. TUTOR gives short feedback on the quality of the answer.

4. TUTOR and STUDENT have a multi-turn dialogue to improve the answer.

5. TUTOR assesses whether the student understands the correct answer.

Notably, step 4 of the frame typically involves leading questions, which the original authors call scaffolding (Graesser, 1993). Leading questions were mentioned by Vygotsky (1978) as a kind of guidance or collaborative support that could be provided by the "knowledgeable other," so this use of the term scaffolding is more in line with Vygotsky than Wood et al. However, in the terminology of Wood et al., asking leading questions is an instance of marking critical features, one of the six scaffolding functions. Step 4 is just one point of clear alignment, and there are additional alignments between some of the observed tutoring behaviors from both novice tutors (Graesser, 1993; Graesser et al., 1995) and expert tutors (Person & Graesser, 2003) in formal educational domains with the six scaffolding functions. We consider these in turn.

As noted by Wood et al., recruitment to the task seemed to be more necessary for 3-year-olds than later ages. Accordingly one would expect little recruitment would be needed for high-school or college students. While there is no clear connection to recruitment with the 5-step tutoring frame, we note that the use of concrete and motivating examples by expert tutors (Person & Graesser, 2003) may serve as a means of recruitment by virtue of being meaningful, authentic, and culturally relevant. Expert tutors average 26 examples per hour (Person & Graesser, 2003), so if examples serve a recruitment function it may be the case that formal educational domains require ongoing recruitment to the task, even with older students. The scaffolding function of reducing degrees of freedom is more clearly present in the 5-step tutoring frame. During step 4, tutors tend to ask students mainly verification questions and concept completion questions (Graesser & Person, 1994). Concept completion questions, sometimes called prompts, typically query a single noun phrase missing in the student's answer. Concept completion questions can be considered a kind of simplification because, instead of asking the student for the complete answer, the tutor asks the student to fill in part of the answer while providing contextual cues in the

question itself to make this task easier. Verification questions may serve to simplify the task to an even greater degree.

Direction maintenance is provided in two ways. It appears that both novice and expert tutors use a curriculum script (Putnam, 1987), a loosely defined ordering of topics, skills, and learning objects that the tutor plans to cover during the tutoring session (Graesser, 1993; Graesser et al., 1995; Person & Graesser, 2003). Expert tutors have also been described as using highlighting, a mode or phase of the tutoring session that redirects the student having trouble with a problem, breaks the problem down, and creates a plan for the student to follow (Cade, Copeland, Person & D'Mello, 2008). Marking critical features appears to encompass both asking leading questions as well as feedback. Leading questions, like hints, typically suggest what the student should be paying attention to or thinking about without providing much additional context, e.g., "What can you say about gravity in this situation?" Feedback may be simple, e.g., "Correct," or elaborated, e.g., "That's correct because gravity acts in a vertical direction."

The fifth function, frustration control, may not be present in more formal educational tutoring contexts. However, several researchers have noticed that both novice and expert tutor feedback tends to be less discriminating than warranted, particularly with regard to incorrect student answers (Graesser et al., 1995; Person & Graesser, 2003). These researchers have proposed that tutors may be trying to maintain student motivation by giving indirect feedback to incorrect answers or by avoiding giving negative feedback at all. Finally, demonstration is evident in novice tutor's splices and summaries (Graesser, 1993; Graesser et al., 1995). Splices are tutor repairs of student's incorrect answers that either partially or completely give the solution to the problem. Summaries are a kind of retrospective demonstration by the tutor, often given when the tutor has decided to move on to another topic. Because they recap the solution steps, they are a kind of demonstration yet distinct from the prospective demonstration described by Wood et al. In expert tutoring, the connection to demonstration is clearer: modeling modes have been identified that are synonymous with demonstration (Cade et al., 2008).

As indicated by the preceding discussion, there does appear to be a correspondence between scaffolding as defined by Wood et al. and naturalistic observations of human tutoring in formal educational domains. This is perhaps surprising because the original context of scaffolding, blocks puzzles, is clearly informal and bears little superficial resemblance to formal educational tasks. Indeed, at a deeper level, there seem to be significant differences in terms of prior knowledge and abstract operations. However, it appears that the scaffolding functions defined by Wood et al. are still relevant. We argue that this is additional evidence that scaffolding is a theoretically well-defined phenomenon rather than a metaphor.

Scaffolding: Implications for Intelligent Tutoring Systems

In this section, we describe later developments in Wood et al.'s theory of scaffolding and how these correspond to theoretical and actual implementations in intelligent tutoring systems. We argue both the original theory of scaffolding and some implementations of ITSs have failed to incorporate parallel developments of scaffolding falling under the umbrella of cognitive apprenticeship (Collins et al., 1991). We further argue that the notion of "making thinking visible," as championed by cognitive apprenticeship, is a significant advancement of scaffolding in formal educational domains where much of the problem space is covert.

It appears that the theory of scaffolding outlined in Wood et al. (1976) remained unchanged in the following decade. Later developments saw a reframing of the notion of scaffolding in terms of contingency (Wood & Wood, 1996a, 1996b; Wood & Wood, 1999; Wood, 2001, 2003). In many ways, this notion of contingency is an elaboration of the Wood et al. (1976)'s notion of a tutor consulting both a student and domain model when making decisions about what to do next. The overarching premise is that tutor

guidance and collaboration should be closely calibrated to the current needs of the student, no more or less. Perhaps the clearest explanation of the theoretical framework of contingency is given in Wood (2003), which defines three dimensions of contingent tutoring:

Instructional contingency: How to support activity

1. General verbal intervention
2. Specific verbal intervention
3. Specific verbal intervention plus nonverbal indicators
4. Prepares for next action
5. Demonstrates action

Domain contingency: What to focus on next (both at the problem and step level)

Temporal contingency: If and when to intervene

Of the three types of contingency, instructional contingency has been the best theoretically developed. In simple terms, it defines a degree of specificity in the tutor's actions, ranging from vague to highly specific. The other two forms of contingency Wood (2003) appears to feel are resistant to further specification, noting that domain contingency is domain-dependent and dynamically grows with the tutor's experience and that temporal contingency is difficult to specify because it is unclear how long one should wait for a student to struggle before offering help.

Although Wood (2003) doesn't explain exactly how the six scaffolding functions of Wood et al. (1976) fit into the notion of contingency, there appear to be a number of correspondences. As discussed earlier, all of the six scaffolding functions may be related to verbal tutor actions in naturalistic tutoring in formal educational domains. Two of the six scaffolding functions, recruitment and frustration control, don't directly relate to instruction in the context of problem solving, however, but rather relate to keeping the student in a state of readiness for instruction. Therefore, it seems that the remaining four scaffolding functions that do relate to instruction in the context of problem solving can each be calibrated according to the specificity of the tutor's action, whether it be reduction in degrees of freedom (the number of degrees reduced), direction maintenance (the specificity of direction given), marking critical features (the specificity with which features are marked), or demonstration (whether a partial, complete, or idealized demonstration is given).

Wood's later focus on contingency seems intermingled with the desire to implement scaffolding and contingency in ITSs (Wood, 2001). One of these tutors, called QUADRATIC, exhibited only instructional contingency, not domain or temporal contingency (Wood & Wood, 1999; Wood, 2001, 2003). All students proceeded through a fixed sequence of problems (eliminating domain contingency) and assistance was only provided in response to student's help requests (eliminating temporal contingency) though the help was calibrated to the individual student (thus having instructional contingency). An evaluation of QUADRATIC revealed that while low-ability students sought help more often that high-ability students overall, high-ability students were more likely to seek help after they made an error than low-ability students. Another tutor, DATA, included both domain contingency and instructional contingency but not temporal contingency (Wood, 2001). Domain contingency was achieved by using a pretest to identify specific categories of errors students were making, and then assigning problems during tutoring based on each child's error pattern. A corresponding evaluation of DATA showed that by including domain contingency, the help-seeking behavior of low- and high-ability students equalized.

We argue that when considered in the context of the original theory of scaffolding, these results are expected: given a set of problems within the ZPD of high-ability students, it is likely that these problems are outside the ZPD of low-ability students because they cannot comprehend the solution, even when it is demonstrated to them. As such, these experiments provide additional support to the original theory of scaffolding. These contingency theory tutors are interesting in another regard: they have conceptual overlap with the ACT* based ITSs as noted by Wood and Wood (1996b). Specifically, the contingency theory and scaffolding overlap with the eight principles for the design of Cognitive Tutors derived from the ACT* theory (Farrell, Anderson, Reiser & Boyle, 1987; Anderson, Corbett, Koedinger & Pelletier, 1995):

1. Represent student competence as a production set

2. Communicate the goal structure underlying the problem solving

3. Provide instruction in the problem-solving context

4. Promote an abstract understanding of the problem-solving knowledge

5. Minimize working memory load

6. Provide immediate feedback on errors

7. Adjust the grain size of instruction with learning

8. Facilitate successive approximations to the target skill.

As noted by Wood and Wood (1996b), Principles 2, 3, 6, and 8 are conceptually linked to contingent tutoring. We briefly elaborate on Wood and Wood's analysis to further show the linkages to the original theory of scaffolding. Taking place within a problem-solving context is what makes contingent tutoring possible. It is not a scaffolding function but is instead a necessary condition for scaffolding to occur. Communicating goal structure is somewhat akin to direction maintenance, though without the affective components. Providing immediate feedback on errors is a means of marking critical features. Finally, facilitating successive approximations to the target skill is not a scaffolding function but rather the contingency at the heart of scaffolding. Successive approximations must imply a reduction in tutor support. In addition to these principles noted by Wood and Wood, we further argue that both minimizing working memory load and adjusting the grain size of instruction parallel the scaffolding function of reducing degrees of freedom.

The comparison with Cognitive Tutors is fruitful because it brings into focus some striking ways in which they differ from contingent tutors, further suggesting avenues of future development for both Cognitive Tutors and contingent tutors. The two ways that contingent tutors differ from Cognitive Tutors are in communicating the goal structure and in promoting abstract understanding of problem-solving knowledge. Above we noted that communicating goal structure was akin to direction maintenance. While this is true, direction maintenance is more aligned with keeping the student focused on a goal, which can happen as a side effect of communicating goal structure. However, communicating goal structure implies much more. The idea behind communicating goal structure is to help the students acquire the ability to decompose goals into subgoals on their own. In the case of learning to write recursive programs (Farrell et al., 1987), the goal structure can be covert. Promoting abstract understanding of problem-solving knowledge is similar in that it invokes the underlying structure rather than surface structure (Anderson et al., 1995). In both cases, the knowledge the student needs is not overt, so extra care is taken to help the student acquire it. It's worth noting that Anderson et al. (1995) implemented communicating goal struc-

ture by providing interface elements like graphs to communicate the structure, and they promoted abstract understanding of problem-solving knowledge through help and error messages. We would argue that these implementations of the principles are rather weak in terms of contingency.

The notion of making the covert overt in the context of scaffolding has been championed in the framework of cognitive apprenticeship (CA) (Collins et al., 1991). A full review of CA is beyond the scope of this chapter, but we briefly remark on its similarities and differences to the original theory of scaffolding. First and foremost, CA casts learning as the development of expertise, with teachers as the experts. Just as in traditional apprenticeship, where the apprentice eventually becomes a master, CA views cognitive apprenticeship as a process by which students can become experts. The first step in enacting CA is to "identify the processes of the task and make them visible to students" (Collins et al., 1991, p. 3). This specifically involves communicating an expert's reasoning and strategies. Expert reasoning and strategies are a crucial missing component to Wood et al. (1976)'s scaffolding. Their scaffolding focused on problems and skills and made no mention of strategies. The omission of reasoning and strategies might have resulted from the domains investigated, which allowed little room for such strategies. Similarly strategies are poorly represented in later work on contingent tutoring. While this work discusses how contingent tutoring may lead students to discover and use strategies by induction, it says nothing about making expert reasoning and strategies visible to students (Wood, 2003). We argue that CA addresses communicating goal structure and abstract understanding of problem-solving knowledge proactively, and proactively addressing these principles in intelligent tutoring systems should yield similar benefits as found elsewhere (Palincsar & Brown, 1984; Schoenfeld, 1985).

Conclusion

In summary, we have analyzed the notion of scaffolding from its genesis to the modern conception of the term. We argue that scaffolding today, which is widely considered a metaphor for support, ignores the most useful and productive theoretical aspects of the original theory. Scaffolding has clear correspondence to and elaborates upon the ZPD. It adds a great deal of clarity to that term by recasting it in terms of the gap between comprehension and production. Scaffolding, though originally defined for small children working on blocks puzzles, has proven to be sufficiently general to accommodate findings from observations of naturalistic tutoring in formal educational domains. However, scaffolding as originally defined may require modeling that is too fined grained for a teacher to implement in a classroom. Further developments in the original theory of scaffolding have added more abstract descriptions of contingency, but these have not replaced the original theory. Indeed the original theory has multiple points of intersection with the principles derived from ACT* that were used to develop the Cognitive Tutors. What is missing from the original theory of scaffolding, and what is also underdeveloped in the principles derived from ACT*, is a proactive approach to supporting the development of expert reasoning and strategy use. CA outlines a framework for making thinking visible in order to support development of expert thinking and strategy use. We argue that scaffolding should move beyond "modeling a solution" to include "modeling a strategy." Only then will scaffolding directly support the learning of hidden goal structures and abstract problem-solving knowledge. In other words, we argue that the most effective scaffolding makes expertise visible.

Acknowledgments

findings and conclusions, or recommendations expressed in this chapter are those of the author and do not represent the views of these sponsoring agencies.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R. & Pelletier, R. (1995). Cognitive tutors: lessons learned. The Journal of the Learning Sciences, 4(2), 167-207.

Belland, B., Walker, A., Olsen, M. W. & Leary, H. (2012, April). Impact of scaffolding characteristics and study quality on learner outcomes in STEM education: A meta-analysis. Paper presented at the 2012 Annual Meeting of the American Educational Research Association.

Brandt-Kobele, OC. & Hohle, B. (2010). What asymmetries within comprehension reveal about asymmetries between comprehension and production: The case of verb inflection in language acquisition. Lingua, 120(8), 1910-1925. http://www.sciencedirect.com/science/article/pii/S0024384110000276

Brown, A. L. & Palincsar, A. S. (1986, March). Guided, cooperative learning and individual knowledge acquisition. Technical Report No. 372.: Bolt, Beranek and Newman, Inc. and Center for the Study of Reading at the University of Illinois Urbana-Champaign. http://www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED270738 (Sponsoring Agency: National Inst. of Child Health and Human Development (NIH), Bethesda, MD)

Bruner, J. S. (1962). Introduction. In E. Hanfmann & G. Vakar (Eds.), Thought and language (pp. 5-10). Cambridge, MA: MIT Press.

Bruner, J. S. (1986). Actual minds, possible worlds, (pp. 70-78). Cambridge, MA: Harvard University Press.

Cade, W. L., Copeland, J. L., Person, N. K. & D'Mello, S. K. (2008). Dialogue modes in expert tutoring. In ITS '08: Proceedings of the 9th International Conference On Intelligent Tutoring Systems (pp. 470-479). Berlin, Heidelberg: Springer-Verlag.

Cazden, C. (1979). Peekaboo as an instructional model: Discourse development at home and at school. Papers and reports on child language development, No. 17. (Papers and Reports on Child Language Development No. 17). Stanford, CA: Stanford University Department of Linguistics.

Chi, M. T. H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. Applied Cognitive Psychology, 10, 33-49.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T. & Hausmann, R. G. (2001). Learning from human tutoring. Cognitive Science, 25(4), 471-533.

Clark, E. V. & Hecht, B. F. (1983). Comprehension, production, and language acquisition. Annual Review of Psychology, 34(1), 325-349. http://www.annualreviews.org/doi/abs/10.1146/annurev.ps.34.020183.001545

Collins, A., Brown, J. S. & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. American Educator, 15(3), 38-46.

Davis, E. A. & Miyake, N. (2004). Explorations of scaffolding in complex classroom systems. Journal of the Learning Sciences, 13(3), 265-272. http://www.tandfonline.com/doi/abs/10.1207/s15327809jls1303_1

Farrell, R. G., Anderson, J. R., Reiser, B. J. & Boyle, C. F. (1987). Cognitive principles in the design of computer tutors. In P. Morris (Ed.), Modeling cognition (pp. 93-133). Chichester: Wiley.

Fodor, J. (1999). Some reflections on L.S. Vygotsky's thought and language. In P. Lloyd & C. Fernyhough (Eds.), Lev Vygotsky: Critical assessments (Vol. 1, pp. 411-420). London: Routledge.

Graesser, A. C. (1993). Questioning mechanisms during tutoring, conversation, and human-computer interaction (ONR Tech Report No. 93-1). Memphis, TN: Memphis State University.

Graesser, A. C. & Person, N. K. (1994). Question asking during tutoring. American Educational Research Journal, 31(1), 104-137.

Graesser, A. C., Person, N. K. & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. Applied Cognitive Psychology, 9, 1-28.

Greenfield, P. M. (1984). A theory of the teacher in the learning activities of everyday life. In B. Rogoff & J. Lave (Eds.), Everyday cognition: Its development in social context (pp. 117-138). Cambridge, MA: Harvard University Press.

Harzing, AW. (2007). Publish or perish. http://www.harzing.com/pop.htm.

Hurley, P. (2011). A concise introduction to logic (11 ed.). Boston, MA: Cengage Learning.

Jones, G. & Ritter, F. E. (1998). Simulating development by modifying architectures. In M. A. Gernsbacher & S. J. Derry (Eds.), Proceedings of the Twentieth Annual Conference of the Cognitive Science Society (pp. 543-548). Mahwah, NJ: Lawrence Erlbaum Associates.

Lajoie, S. (2005). Extending the scaffolding metaphor. Instructional Science, 33(5-6), 541-557.

Palincsar, A. S. (1998). Keeping the metaphor of scaffolding fresh: A response to C. Addison Stone's "The metaphor of scaffolding: Its utility for the field of learning disabilities." Journal of Learning Disabilities, 31(4), 370-373.

Palincsar, A. S. & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. Cognition and Instruction, 1, 117-175.

Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. The Journal of the Learning Sciences, 13(3), 423-451. http://www.tandfonline.com/doi/pdf/10.1207/s15327809jls1303_6

Person, N. K. & Graesser, A. C. (2003). Fourteen facts about human tutoring: Food for thought for ITS developers. In AIED 2003 Workshop Proceedings on Tutorial Dialogue Systems: With a view toward the classroom (pp. 335-344).

Puntambekar, S. & Hubscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? Educational Psychologist, 40(1), 1-12. http://www.tandfonline.com/doi/abs/10.1207/s15326985ep4001_1

Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. American Educational Research Journal, 24(1), 13-48.

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G. Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. Journal of the Learning Sciences, 13(3), 337-386. http://www.tandfonline.com/doi/abs/10.1207/s15327809jls1303_4

Schoenfeld, A. H. (1985). Mathematical problem solving. Orlando, FL: Academic Press.

Sherin, B., Reiser, B. J. & Edelson, D. (2004). Scaffolding analysis: Extending the scaffolding metaphor to learning artifacts. The Journal of the Learning Sciences, 13(3), 387-421. http://www.tandfonline.com/doi/abs/10.1207/s15327809jls1303_5

Stone, C. A. (1998a). The metaphor of scaffolding and its utility for the field of learning disabilities. Journal of learning disabilities, 31(4), 344-364.

Stone, C. A. (1998b). Should we salvage the scaffolding metaphor? Journal of Learning Disabilities, 31(4), 409-413.

van de Pol, J., Volman, M. & Beishuizen, J. (2010). Scaffolding in teacher-student interaction: A decade of research. Educational Psychology Review, 22, 271-296. http://link.springer.com/article/10.1007/s10648-010-9127-6

Vygotsky, L. S. (1962). Thought and language (E. Hanfmann & G. Vakar, Eds.). Cambridge, MA: MIT Press.

Vygotsky, L. S. (1978). Mind in society. Cambridge, MA: Harvard University Press.

Wood, D. (2001). Scaffolding, contingent tutoring, and computer-supported learning. International Journal of Artificial Intelligence in Education, 12, 280-292.

Wood, D. (2003). The why? what? when? and how? of tutoring: The development of helping and tutoring skills in children. Literacy Teaching and Learning: An International Journal of Early Reading and Writing, 7(1-2), 1-30.

Wood, D., Bruner, J. & Ross, G. (1976). The role of tutoring in problem solving. Journal of child psychology and psychiatry, 17(2), 89-100.

Wood, D. & Middleton, D. (1975). A study of assisted problem-solving. British Journal of Psychology, 66(2), 181-191.

Wood, D. & Wood, H. (1996a). Contingency in tutoring and learning. Learning and Instruction, 6(4), 391-397. http://www.sciencedirect.com/science/article/pii/S0959475296000230

Wood, D. & Wood, H. (1996b). Vygotsky, tutoring and learning. Oxford Review of Education, 22(1), 5-16. http://www.tandfonline.com/doi/abs/10.1080/0305498960220101

Wood, H. & Wood, D. (1999). Help seeking, learning and contingent tutoring. Computers & Education, 33(2–3), 153-169. http://www.sciencedirect.com/science/article/pii/S0360131599000305

Woolf, B. P. (2008). Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning. Morgan Kaufmann Publishers/Elsevier.

EPILOGUE

CHAPTER 27 – Cost Analysis for Training and Education Systems

J.D. Fletcher[1] and Robert A. Sottilare[2]
Science & Technology Division - Institute for Defense Analyses[1]
U.S. Army Research Laboratory[2]

Introduction

Assessments in research and development are intended to inform decisions about theoretical alternatives and/or recommended courses of action. In training and education, these assessments typically focus on the effectiveness of candidate techniques, applications, and programs in producing learning objectives. Such assessments are essential, but for decision makers they are only half the story. Managers and administrators must consider not just effectiveness, but the costs, defined in both monetary and operational terms, of any proposed course of action. Although infrequently considered by education and training researchers, cost factors can be as critical as effectiveness in improving the state of art and practice in training and education. This chapter provides an overview of costs, cost-benefit, cost-effectiveness, return on investment, and net-present value analyses that might be applied in making a business case for advancements and opportunities provided by research in training and education and as factors in research and development for GIFT (Sottilare, Brawner, Goldberg & Holden, 2012).

Why Cost Analysis?

Empirical assessments of training and education programs focus on effectiveness with good reason. The ability of a program to produce learning is a core concern for learners and those providing the program. Much, therefore, has been done to hone and apply our abilities to assess program effectiveness. Research in ITSs, as in other training and educational systems, has focused heavily on effectiveness and their effect sizes in comparisons with the "traditional classroom instructional model." However, the hallmark of decision making, which determines what programs and approaches are actually adopted, is not simply effectiveness, but what must be given up to get it – in short, its cost.

Even though we have conducted over four decades of ITS research and shown ITSs to be effective tools in providing one-to-one-tutoring, ITSs are not ubiquitous. In large measure, the skills needed to author an ITS and the cost of those skills limit ITS use. Even a simple ITS providing one hour of instruction may take 200 hours to develop at a cost of $50,000. This cost may not be practical for low density or low throughput courses.

Cost analysis concerns the resources and opportunities that must be sacrificed to take any course of action along with the value, benefit, or utility it provides. Costs and benefits are usually measured in monetary terms, but measures of productivity, operational effectiveness, health, quality of life, morale, and human life are also fair game as variables of interest.

However, program costs and cost analyses are rarely included in assessments of training and education alternatives. Their absence is unfortunate. If researchers and developers intend the products of their efforts and findings to be used, analysis of costs, undertaken with the same care and professional attention now given to effectiveness, seems essential. The need for cost analysis to inform administrative decisions seems critical when new programs and procedures are being considered as replacements for those already in place.

This need is especially evident in the case of training, which in industry and the military, is not an end in itself but an alternative that must compete with other means used to secure necessary levels of operational productivity and effectiveness. Training must hold its own in competition with other alternatives such as the purchase of materiel (e.g., new equipment and machines) and supplies (e.g., spare parts and consumables). Even within personnel systems, training must compete with alternatives such as personnel selection, ergonomic design, and provision of job aids in producing necessary levels of human performance.

Most decision makers are willing to allocate resources to training and education on faith, but faith can only go so far in competition with alternatives that can and usually do argue their case in terms of both effectiveness and costs. Researchers and developers in the training and educational domains must be able to do the same. They must be able to say what a pound of training is worth.

This chapter provides some background and practical advice to help training and education researchers and developers address issues of cost. It is introductory and not the final word, but it may suggest ways to begin traversing this challenging landscape.

Cost Analysis

Cost analysis serves as an overarching, inclusive term for activities such as efficiency assessment, cost-benefit analysis, return-on-investment, economic analysis, cost-effectiveness analysis, and cost-utility analysis. By assessing both how much is gained and how much must be given up, cost analysis determines the anticipated net value of decisions.

Cost analysis has a long history. As early as 1667, public health officials in London defended their efforts to combat the plague with a benefits to cost ratio of 84:1 (Thompson, 1980). Requirements and mandates for cost analysis continue. The Planning, Programming, Budgeting System was specifically implemented for the Great Society programs of the 1960s and remains a core requirement for defending today's public expenditures with cost analysis. The 1973 "Principles and Standards for Planning Water and Related Land Resources" required cost-benefit trade-offs among economic development, regional economic development, environmental quality, and social well-being in the United States (Sassone & Schaffer, 1978). The continuing emphasis on accountability and on determining quantitative relationships between public investments and their returns suggests a persistent need for cost analysis in training and education.

Current interest in and emphasis on cost analysis began in the manufacturing sector and has now worked its way through service, health care, and public sectors. It is beginning to find application in industrial training (Phillips, 2011), military training (Cohn & Fletcher, 2010; Fletcher & Chatham, 2010; Orlanksy & String, 1979, 1981), and education (Levin & McEwan, 2001), but it has yet to be established as an essential and routine component of assessment in either training or education – a matter lamented by researchers such as Rice (1997), Hummel-Rossi & Ashdown (2002), Ross, Barkaoui & Scott (2007), and Harris (2009). The need for cost analysis in training and education is growing as alternative investments increasingly defend their budgets and expenditures with cost analysis. Rice (1997) described the neglect of cost in training and education assessments as a paradox, wondering "why such a seemingly relevant form of analysis has been so underutilized" (page 309). Ross, Barkaoui, and Scott (2007) note that the neglect of cost analysis remains unabated.

Cost analysis is as subject to controversy as any other assessment. Differences in approaches and procedures are as likely to be found in cost analysis as elsewhere. Decision makers in training and education would be well served by the adoption of generally accepted, standardized cost element definitions, data models, and analysis techniques so that they and others understand what they are talking about. Standards

and specifications have been suggested (e.g., Knapp & Orlansky, 1983) but none are commonly accepted and used.

Beyond these methodological issues, necessary cost data may only approximate the needs and objectives of an analysis – guesses and projections are not unusual. Further, an analysis, with its data and findings, is intended to inform a specific decision or class of decisions. Decision makers with different but similar concerns may consult it later for their own purposes.

The real-world consequence of these conditions is that cost analyses are rarely, if ever, perfect. However, they can, and should be compulsively explicit and documented so that reviewers know what was done, how, and why. Approximations, models, assumptions, and cost element definitions should be clearly reported, enabling decision makers to decide for themselves if and to what extent a cost analysis may be used to inform their decisions.

Cost Categories

To develop a model or at least a framework for cost analysis, we need to consider the categories of costs that are commonly included. Costs generally fall into four major categories: *Research and Development*; *Initial Investment*; *Operations and Maintenance*; and *Disposal and Salvage* (Mishan & Quah, 1988). Research and Development costs account for the materials, people, and facilities needed to create and evaluate a new program or capability. Initial Investment costs cover the one-time costs of procuring and deploying resources to implement it. Operations and Maintenance cover costs needed to manage, operate, support, and maintain a program once it is implemented. Disposal and Salvage costs are the one-time "clean-up" expenses of removing the program from operational use. The sum of these costs is sometimes referred to as life-cycle cost (LCC) or total ownership cost (United States Department of Defense, 1992).

Research and Development and Initial Investment costs for a proposed program are often included in a analysis of total ownership costs, but such costs for the program currently in place are generally omitted. In most cases, nothing can be done to change them, and they are considered "sunk." A new approach or program must then present its net value with the added burden of the Research and Development and Initial Implementation costs it will incur. Disposal and Salvage costs may be considered for candidate and/or existing programs, but, with some exceptions such as the costs of removing existing simulators or disposing of training ranges, they are minor in training and education applications and generally excluded.

Cost Models

A cost model is the foundation for any cost analysis. It identifies, lists, and defines the cost elements that will be included in any analysis that might be performed. As suggested above, specificity and explicitness are critically important for cost analyses. It is as important for analysts to know and articulate what they are talking about as it is for decision makers to understand them. What is not included in a cost element should be as clear as what is included.

Early on, Levin (1983) suggested five classes of elements, or "ingredients," to be considered in a cost model. They are: *Personnel, Facilities, Equipment and Materials, Other Program Inputs, and Client Inputs*.

Personnel costs include all the resources required for the human resources needed by the approach. Levin recommended that personnel be classified according to their roles (instructional, administration, clerical, etc.), qualifications (training, experience, specialized skill), and time commitments (full time, part time). *Facilities* costs include all resources required to provide physical space for the approach. *Equipment and*

Materials include furnishings, instructional equipment, and consumables. *Other Direct Costs* in Levin's scheme include components that do not fit elsewhere, for instance, instructor training and insurance costs.

Client Inputs include resources that must be contributed by students and/or their employers. These inputs are especially relevant in military and industrial training where student salaries and allowances are paid by the client. In these cases, the client has a strong interest in the speed with which learners achieve instructional objectives. An early rationale for applying technology in industrial and military training was keyed to its capacities for self-pacing, acceleration of learning, and earlier release of students for duty (Fletcher, 2009). In effect, this approach was to decrease costs (time to learn in this case) while holding effectiveness (amount learned) constant rather than to increase effectiveness (learning) while holding costs (time) constant. There is more to be said about this distinction and associated practicalities later.

About the same time, Kearsley (1982) developed a model with one dimension very much like Levin's but with an added dimension keyed to the ADDIE stages of instruction system development – the familiar system engineering stages of *Analysis*, *Design*, *Development*, *Implementation*, and *Evaluation*. Although Kearsley's approach is intended for use in training, its cost elements are readily adapted for education.

By replacing *Materials* with *Consumables*, bundling *Client Inputs* with *Other Direct Costs*, and adding a row for *Indirect Costs*, the framework suggested by combining Levin's and Kearsley's approaches is as shown in Table 1. *Indirect costs* are those over which an activity has little or no direct control. For example, they include its share of operating costs (e.g., electricity, road maintenance, security, rents) for the site (e.g., school grounds, military base, business park) where the activity is physically located and the operating costs of its parent organization (e.g., university, corporation, government agency). They may be identified as overhead or administrative expenses, and they may be fixed or variable. These costs are usually prorated among the activities at a site. Discussions over what are and are not indirect costs and their fair apportionment are not uncommon.

Table 1. A Cost Model Framework for Training and Education.

	Analysis	**Design**	**Development**	**Implementation**	**Evaluation**
Personnel					
Facilities					
Equipment					
Consumables					
Other Direct Costs					
Indirect Costs					

Example cost components for the rows in Table 1 may be familiar to anyone developing project budgets. Some examples are shown in Table 2. Knapp and Orlansky (1983) developed a more complete set of costs for military training that lists 75 elements in all. Some of these elements can be eliminated in estimating the costs of training and education alternatives because they are the same across all alternatives, irrelevant, or sunk. Development of any cost model should be keyed to the decision(s) it is intended to inform.

Table 2. Example cost components.

Personnel	Facilities	Equipment	Consumables	Other Direct	Indirect
Managers & Administrators	Rents	Production & tooling hardware	Spare parts	Instructor training	Site administration & management
Security personnel	Mortgage & interest	Computer hardware, software	Fuel and lubricants	Learner salaries & allowances	Rents
Clerical & Library personnel	Utilities		Ammunition		Mortgage & interest
	Training ranges	Other instructional equipment	Office materials	Transportation, lodging, and per diem costs	
Analysts & evaluators	Kitchens	Instrumentation	Materials for simulated environments		Utilities
	Storage	Office equipment			Range support
Subject matter experts	Maintenance & cleaning		Assessment & data collection materials		Roads & parking
		Vehicles			Transportation
Instruction designers	Data collection & storage	Simulators			Maintenance & cleaning
Instructors		Facility support equipment			Security
Technicians					Food service
Implementation coordinators					
Consultants					

Types of Analysis

Varieties of cost analysis include cost-benefit, cost-effectiveness, cost-utility, return on investment, and net present value. These are all based on ratio calculations. Here and elsewhere, cost analysis serves as an overarching, inclusive term for all of them.

Cost and Benefit Analysis

Cost and benefit analysis determines how much the benefits or value returned by some course of action outweigh, if at all, its costs. It is generally expressed as the ratio of benefits to costs. We can calculate a benefit/cost ratio using whatever metrics we choose, but they must be commensurable – both benefit and cost must be assessed with same unit of measure. The usual, but not exclusive, way of dealing with this constraint is to reduce all costs and benefits to monetary units – dollars, euros, yen, etc.

As discussed by Phillips (2011), among others, a benefit/cost ratio is calculated as

$$\text{Benefit/Cost} = \frac{\text{Value of the Result}}{\text{Cost of the Result}}$$

It tells us, in quantified terms, how many units of value we get for every unit of cost. Typically, we seek a value greater than 1.0. An example of a cost and benefit analysis involving a GIFT issue follows.

It has been determined that an "intelligent" authoring tool to extract expert models from text-based sources will result in a 50% reduction in development time for expert models in well-defined domains (e.g., mathematics, physics, or technical training). The cost of developing the authoring tool and deploying the tool within GIFT is $1.5 million. If the average development cost for an expert model for a one-hour course is $20,000 today and an average of 4000 hours of instruction is produced each year, what is the benefit/cost ratio in the first year for this domain?

$$Value \ of \ the \ Result = 0.50 \ (\$20,000 * 4000) = \$40M$$

$$Benefit/Cost = \frac{40}{1.5} = 26.67$$

In this example, then, an investment of $1.5M will produce savings of $40M over the first year of its implementation, yielding a benefit to cost ratio of 26.67 – assuming everything else is equal.

Return on Investment (ROI)

ROI may be viewed as a net benefit to cost ratio. Again, as discussed by Phillips (2011) and others, it is

$$ROI = \frac{Value \ of \ the \ Result - Cost \ of \ the \ Result}{Cost \ of \ the \ Result}$$

This ratio is often multiplied by 100 so it can be expressed as a percent rather than a proportion. Any return on investment greater than 100% (when all costs are recovered) indicates a positive net return. Return on investment must be calculated for a specified period of time, such as a year. As with monetary units, the length of time covered should be determined by analysts in consultation with decision-makers who are likely to use the results of the analysis.

Using values obtained from the previous benefit/cost example, we have

$$ROI = \frac{40 - 1.5}{1.5} = \frac{38.5}{1.5} = 25.67$$

Like cost and benefit analysis, return on investment requires value and cost to be commensurable, or calculated using the same basis of measurement – in this case, dollars. Of the two, return on investment may be preferred because it indicates how many units of net benefits are returned (after investment costs have been subtracted out) per unit of cost. There are, of course, spikes, dips, and diminishing returns to be considered with differently timed units of investment, so averaging and curve smoothing may be required.

The issues that arise with commensurability and return on investment usually concern what cost model should be used, how its cost elements should be defined, and what values should be assigned to parameters such as discount, interest, depreciation, inflation, and amortization rates. These assignments often involve assumptions, extrapolations, and best guesses. This is especially the case when returns are sought in terms of human performance. For these reasons, we may turn to cost-effectiveness analysis.

Cost-Effectiveness Analysis

In cost-effectiveness analysis, the costs of investment are generally expressed in monetary units. However, benefits can be expressed in non-monetary terms such as human performance (e.g., accuracy of troubleshooting and repair, number of patents), the operational capability of a military unit (e.g., sortie rates, communication efficiency), or business productivity (e.g., sales, product delivery). Cost-effectiveness analysis allows returns to be measured in their own units and incorporated in the analysis. It allows the analysis to cover a more complete range of outcomes.

Ratios are (again) used to express cost-effectiveness. A cost-effectiveness ratio reports either effectiveness produced per unit of cost or, the reciprocal, cost per unit of effectiveness – either effectiveness or cost is held constant leaving the remaining factor free to vary across all alternatives under consideration.

Cost-effectiveness is a relative term, and relevant decision alternatives must be specified in its assessment. Despite common usage, we cannot properly say an investment, simply by itself, is or is not cost-effective, although a cost-effectiveness ratio may legitimately be calculated for it. For example, Fletcher, Hawley, and Piele (1990) held effectiveness constant and calculated costs of different education interventions (decreasing class size, lengthening the school day, providing professional tutors, providing peer tutors, and using computer-assisted instruction) required to raise mathematics scores one standard deviation on a standard test of mathematics. Additional examples of this sort and more extensive discussion are presented by Ross, Barkaoui & Scott (2007).

Possible outcomes of a cost-effectiveness analysis are shown in Table 3. If an alternative reduces costs to produce a fixed outcome or if it increases the outcome value without increasing costs, it presents a strong case for its adoption. Ambiguity arises when higher cost also increases the value of an outcome, or lower cost also reduces the value of an outcome. Choices in these cases may key on the magnitude of the effects, or the urgency of increasing the outcome productivity or value. However, fungible monetary units may be, they may diminish in importance the closer the choice of an alternative is, for instance, to affecting the outcomes of a military engagement.

Table 3. Decision space for cost-effectiveness assessment.

		COSTS	
		Increase	Decrease
EFFECTIVENESS	Increase	??	Accept
	Decrease	Reject	??

Choice of alternatives is a critical issue in cost-effectiveness analyses. The addition or modification of an alternative after a cost-effectiveness analysis is finished may well affect its conclusions and recommendations. If a single course of action is to be selected based on cost-effectiveness analysis, the set of alternatives should be as well defined and comprehensive as possible. An additional step to take is to parameterize components of the analysis so that different values may be readily substituted to determine their effect on the results. This step may be taken as a matter of course during a sensitivity analysis, which is briefly discussed later.

For example, a school district is considering adopting an ITS to support remedial reading. Within this school district approximately 22% of its 10,000 students in grades 1–6 require additional reading instruc-

tion and practice each year to meet district standards. This tutoring is currently outsourced to independent (human) tutors within the district at an annual cost of $4.3M. The resulting effect size is 0.8 standard deviations when compared to traditional classroom instruction. The costs of implementing three different ITSs (A, B, and C) are $2.4M, 3.6M, and 5M with effect sizes of 0.50, 1.05, and 1.30, respectively. Based on cost-effectiveness and assuming that these effect sizes apply equally for 2200 students in grades 1–6, which method should we chose?

Results from examining each alternative, are shown in Table 4.

Table 4. A simple cost-effectiveness example.

Tutoring Alternatives	Cost ($M)	Effect Size	Cost per Effect Size Unit
Human	4.3	0.80	4.3/0.80 = 5.38
ITS A	2.4	0.50	2.4/0.50 = 4.80
ITS B	3.6	1.05	3.6/1.05 = 3.43
ITS C	5.0	1.30	5.0/1.30 = 3.85

If we want an improvement at the least possible cost, we would choose ITS A. If we want to maximize the increase in reading ability regardless of cost, we would choose ITS C If we want to increase reading ability in the most cost-effective manner, we would choose ITS B. The choice is up to the decision maker, but the analysis (in this simple example) is a factor that should help inform the decision.

This example makes a number of assumptions, for instance that the effect sizes are accurate at this scale (2200 students), that the cost estimates are realistic, that either the effect sizes will be nearly the same in all six grades for all teachers or the reading instruction will be provided to an equal number of students in each classroom in each grade, and so forth. This example was to just provide an illustration of cost-effectiveness analysis. In the real world, issues such as these would have to be accounted for, but they should be familiar to researchers experienced in field work.

Cost Analysis Over Time

Development of training and education systems often involves a series of investments made over several time periods (e.g., years). They require a time-series approach if they are to be included in an analysis intended to assess the costs and benefits of introducing an alternative instructional system. Such an analysis generally requires inclusion and assumptions of *inflation*, *discount rates*, and *depreciation*. All three factors are projections. As with all projections and assumptions, the cost model and its results should enable decision makers to assess sensitivity to variations in these factors.

Inflation adjustments account for changes in the level of prices for investments and returns made over a period of years. They are usually keyed to the year the investment is made and allow reporting in constant monetary units from that year. Inflation adjustments use one of several available price indexes. The analysis should document which index is used and why.

Discounting is different from inflation. Discount rates are intended to account in the present for expenditures to be made in the future. They account for the possibility that funds expended in a series of investments over time are denied the returns they might receive if invested elsewhere. Like inflation, the discount rate to apply each time period, is a matter of concern and debate. As with inflation, the analysis should document which discount rates were used and why.

The U.S. Office of Management and Budget (OMB) updates both "nominal" (market) and "real" (inflation-corrected) discount rates annually and publishes them in Appendix C of OMB Circular A-94. Nominal discount rates are directly based on the projected prices of goods and services. Real discount rates concern the costs of goods and services relative to those of other or other-year goods and services.

OMB rates are widely used, but different rates are available from other sources such as the Congressional Budget Office and the General Accounting Office. Levin and McEwan (2001) recommend beginning with rates of three to five percent, which may suffice for many purposes.

Depreciation accounts for the decrease in value of assets, such as physical facilities and equipment, over time. Depreciation is used in cost analysis to assess 1) the cost to replace an asset or do without it, 2) its useful life span, and 3) the costs of continuing its use rather than investing that support elsewhere. A related factor, amortization, estimates the useful life span of an asset.

Net Present Value (NPV)

NPV estimates how much in net payoff an investment is worth at the time it is made. It is increasingly used in cost analysis to analyze potential investments. With few exceptions, NPV is mandated for assessments of federal programs. It is defined as the difference of discounted benefits less discounted costs. It provides a way to track the effectiveness of investments over the acquisition lifecycle – including, if needed, early periods of research and development if the results are not already sunk.

NPV analysis can be used by itself to determine if a proposed effort is viable, but it can also be applied in cost-effectiveness assessments. If, for instance, we wish to compare the cost in present day funds of developing and implementing a new instructional approach versus continuing with the approach already in place, we could calculate the NPV for each over a number of years using a discount rate. The preferred approach would be the one with the greater NPV.

NPV may be calculated as

$$NPV = (B_1 - C_1) + \frac{(B_2 - C_2)}{(1+i)^{t-1}} + ... + \frac{(B_n - C_n)}{(1+i)^{t-1}} = \sum_{t=1}^{n} \frac{B_t - C_t}{(1=i)^{t-1}}$$

where

- B_t is the value of the benefit associated with the investment in year t.

- C_t is the cost associated with the investment in year t.

- n is the number of years in which the investment generates benefit and/or costs.

- i is the discount, or interest, rate for year t.

Going back to our original example, what would be the NPV of investment in an ITS for remedial reading over three years assuming we chose ITS B – whose initial implementation cost was $3.60M, an annual operation and maintenance cost of $1.20M, annual savings of $4.30M, which were no longer being spent annually for the human tutoring, and a discount rate of 4 percent per year?

These assumptions give us

$$B_1 = 4.3, B_2 = 4.3, \text{ and } B_3 = 4.3$$

and

$$C_1 = (3.6 + 1.2), C_2 = 1.2, C_3 = 1.2$$

$$NPV = (4.3 - 4.8) + \frac{4.3 - 1.2}{(1 - 0.04)^1} - \frac{4.3 - 1.2}{(1 - 0.04)^2} = 5.35$$

Without NPV discounting, the value over the three years would amount to \$5.70M, but applying it in years 2 and 3, when initial year dollars are worth less, today's NPV would be \$5.35M. That would represent a 3-year difference of \$350,000 with discounting. In any case, the \$5.35M cost of the program would be far less than the (3 * 4.3) = \$12.70M in NPV if current use of human tutors were continued and the annual cost of that program remained unchanged.

A real-world example was provided by Cohn and Fletcher (2010), who applied NPV discounting to compare the costs of developing and implementing a digital tutor that, in 16 weeks of residential training, accelerated the development of information technology expertise beyond that of Navy technicians averaging more than seven years of on-job-training.

Cost Analysis for Simulation

Classical notions of system simulation (e.g., Shannon, 1975) involve models of real equipment, processes, and organizations that can be manipulated for training and experimentation. The expectation that these models can be manipulated leads to the idea of interactive, dynamic simulations that are fundamental to GIFT and other problem-based instruction (Sottilare, Brawner, Goldberg & Holden, 2012; Towne, 1995). These simulations furnish interactive learning environments with safety, visibility, time control, and especially the economies gained from use of virtual material, equipment, and situations in place of the real thing. Their effectiveness in providing knowledge and skills that successfully and, more to the point, affordability transfer to "real" environment is central in their assessment. Cost analysis seems a natural, if not essential, component in the use of simulation for training and education.

Training for operation and maintenance can involve simulated systems that range from simple tools to complex assemblies such as radar repeaters, automotive electronics, and nuclear power plants. Simulations for higher level decision-making may concern issues such as inventory control, emergency management, and command and control. As specified by GIFT, they must provide sufficiently detailed descriptions of their internal structures and their operation to be used in meeting conceptual and other higher-order objectives in training and education. At some point in an instructional program using simulation, learners must turn from simulation and transfer their knowledge and skills to operations in the real environment. Assuming that simulation or simulator time is less expensive than time in the real environment, analysis becomes a matter of maximizing simulator time, minimizing real equipment time, and factoring in the costs of both.

A prototypical example for such analysis is the use of simulators for training aviators. Roscoe and Williges (1980) addressed this issue in general with transfer effectiveness ratios (TERs), which they defined as

$$TER = \frac{T_C - T_X}{X}$$

where

- T_C is time or trails for a control group (e.g., flying the aircraft) to reach criterion performance.

- T_X is time or trails for an experimental group using the control group approach (e.g., flying the aircraft) to reach criterion performance after using the experimental approach (e.g., the simulation).

- X is time or trails spent by an experimental group using the experimental approach (e.g., simulation).

In effect, a TER determines how much net experience (e.g., time, trials) with any real system is saved for every unit of simulation usage invested. Cost analysis using TER then keys on the relative costs of the simulation and, in this instance, the aircraft. If the TER finds that six hours in the simulation save one hour in the aircraft, simulation training is fair game as long as the costs of operating the simulation are less than one-sixth the costs of operating the aircraft.

At least three cautions are in order considering the use of TERs in cost analysis.

First, not all simulation training hours are equal – early hours in simulation may save more real-system time than later ones – the curve is usually negatively accelerated and the returns diminish over time in simulation. This consideration leads to learning curve differences measured by incremental transfer effectiveness.

Roscoe and Williges (1980) defined incremental transfer effectiveness ratios (ITERs) ratios as

$$TER = \frac{T_{X-\Delta X} - T_X}{\Delta X}$$

where

- X is time or trials in simulation.

- ΔX is time or trials in simulation after completing X-ΔX time or trials.

- T_X is time or trails for an experimental group using the control group approach (e.g., flying the aircraft) to reach criterion performance after using the experimental approach (e.g., the simulation).

- $T_{X-\Delta X}$ is time or trials in the control condition to reach criterion performance after completing x-Δx time or trials in simulation.

As an operational example, Taylor et al. (2002) used ITER to determine the point at which flight instrument training hosted by a personal computer was no longer of value. Their study indicated that 5 hours using the computer for this training was cost-effective, but that more simulator (personal computer) time was not. Their results support a policy that permits personal computer time for training instrument tasks in lieu of time in an approved training device or airplane, but they also suggest that only 5 of those hours can be used cost-effectively.

Second, a key argument for simulation is that knowledge and skill acquired in a simulated environment will transfer to the real world. However, transfer may be tied to the acquisition of specific skills or achievement levels. Estimates of transfer from a training simulation must consider what objectives are targeted by the training. Holman (1979) examined this issue and found that if the training transfer was simply focused on general ability to fly a helicopter, the overall TER was 0.72. However, he also found that TERs for 24 specific skills required to fly the helicopter ranged from 2.8 to 0.0. As in all assessments, which TER is relevant depends on the decision it is intended to inform.

The third issue arises from the difficulty of estimating transfer from different amounts of simulation-based training on the time required to reach criterion performance in the aircraft, or whatever the training objectives are targeting. Sufficient estimates might be obtained by reviewing data already available – time required without simulation training and the time required by current simulation-based training applications. It may also be possible, as Morrison and Holding (1990) suggest, using an example from tank gunnery training, to augment available data with limited empirical assessments extrapolated from greater and lesser amounts of simulation-based training combined with estimates from experienced trainers and performers.

These data may be integrated to develop isoperformance curves (Figure 1) that hold performance levels constant with different combinations of simulation and actual equipment training. The cost assumptions underlying such analyses are that 1) time in the simulation costs less than time using the real equipment; 2) more time in the simulation reduces time required to reach criterion in the real equipment thereby reducing overall training costs (with constant levels of performance); and 3) there is a point of diminishing returns, suggested by the upper curve in Figure 3, where so much simulation time is required that overall training costs could even exceed pre-simulation levels (Jones & Kennedy, 1996). Figure 1, adapted from Morrison and Holding (1990), provides a visual representation of isoperformance analysis. These analyses all hold performance constant while seeking the minimal costs to produce it. Carter and Trollip (1980) used a mathematically equivalent approach to devise the opposite – a strategy for maximizing performance (or effectiveness) while holding costs constant.
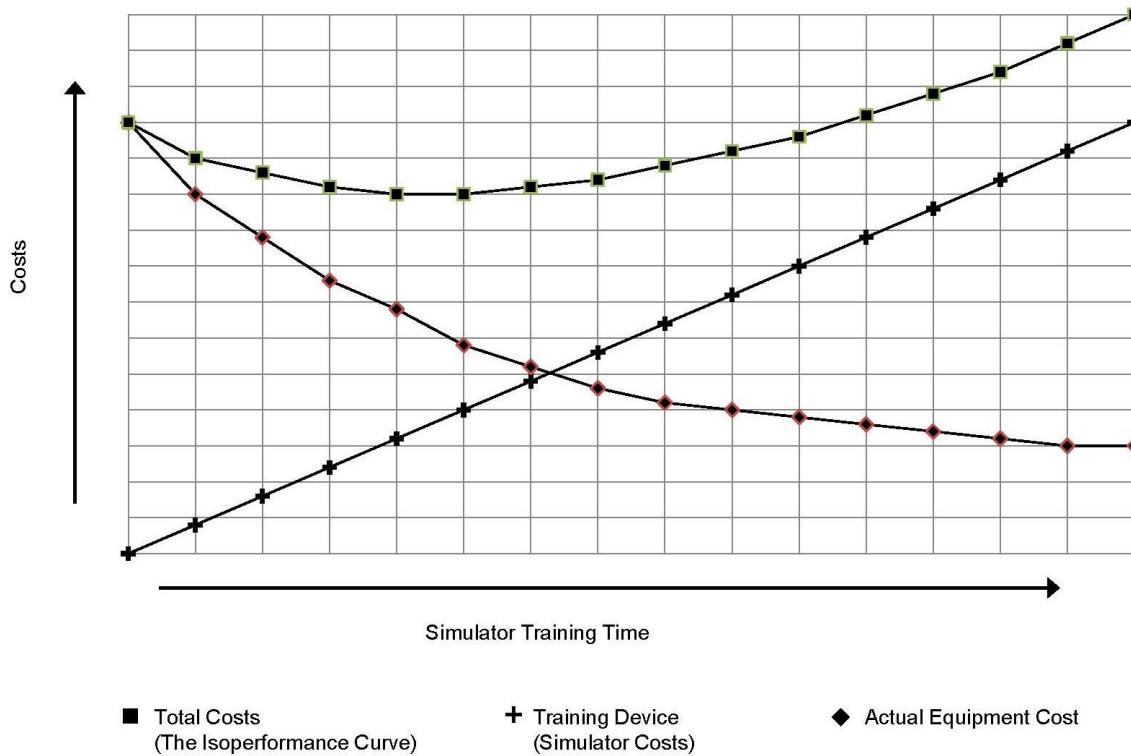
Figure 3. A notional isoperformance curve drawn as a function of simulation and actual equipment costs (adapted from Morrison and Holding, 1990).

Sensitivity Analysis

The values assigned to cost elements may vary with time, technology, decision space, etc. The impact of these variations on analysis findings can and should be assessed by sensitivity analyses, which indicate how robust the findings are when different values are assigned to cost elements. Sensitivity analysis is facilitated when the analysis is presented as a model with parameterized elements so that a change in any parameter will be propagated throughout the model. With or without such facilitation, the selection of parameters to modify in assessing the robustness of analysis findings may be prioritized by the uncertainty with which values were first assigned and the likelihood that altering them will affect analysis findings. All elements in the analysis are fair game, including discount and inflation rates. Values assigned to the elements should be realistic, but the sensitivity analysis should consider both maximum and minimum possibilities.

Cautions, Pitfalls, and Fallacies

A number of commentators list errors or 'fallacies' to be avoided in cost analysis. They might be summarized as the following:

- The *ratio fallacy* is to focus on ratios alone and neglect the magnitude of their cost elements. These magnitudes and/or the risk they imply may be far from the scale decision makers have in mind or can tolerate (e.g., tens vs. thousands of learners, days vs. months for instruction, thousands vs. millions of dollars).

- The *infallibility fallacy* is to assume that a cost analysis supplies the final answer for any decision to be made. An analysis may be essential, but it is based on assumptions and extrapolations that might well change under further scrutiny. It should not be the only factor in the final decision.

- The *extrapolation fallacy* is to neglect the limits of a particular cost analysis and seek its implementation in areas or to a degree beyond those anticipated or considered. An analysis should explicitly list its limitations and decision makers should take them into account.

- The *interrelationship fallacy* is to assume that the relationship between analysis variables is monotonic or even linear. Discontinuous and non-linear relationships may occur in any analysis – especially in those involving human learning and performance. Further, elements that are unrecognized or neglected by the analysis may affect the relationship of variables to one another and the consequences of any action the analysis suggests.

- *The interaction fallacy* is to neglect the effects an instructional program or approach may have on other applications present or planned and, in return, their effect the applications might have on the program or approach the analysis is assessing.

- The *sunk-cost fallacy* is to include cost(s) in the analysis that will not be affected by whatever decision is made.

- The *free-range fallacy* is to leave unspecified the decision(s) the analysis is intended to inform. As belabored in this chapter, no cost analysis is likely to be perfect. Its purpose and its underlying assumptions should be identified and made as explicit as possible.

- The *inflation fallacy* is to ignore the general nature of price indexes and discount rates. Specific costs involved in a decision may or may not change over time in accord with either. These effects should be noted when possible and assessed in sensitivity analysis.

- The *cost-significance fallacy* is to neglect the uncertainty in cost analysis about what constitutes a significant difference. Different costs will have different impact and significance depending on context and scale.

- The *insensitivity fallacy* neglects sensitivity analysis, or the provision of a parameterized model for others to use in performing such an analysis. When the urgency under which the analysis is performed is too great to allow time or resources for follow-on sensitivity analysis, that limitation should be explicitly noted.

GIFT and Cost Analysis

The modular structure of the GIFT architecture supports multiple versions of components and methods within ITS learner, pedagogical, and domain modules. For example, GIFT currently has a default instructional engine known as the engine for Managing Adaptive Pedagogy (eMAP) within its pedagogical module. This engine could be readily replaced with another based on a comparative analysis of maintenance costs and total ownership costs with learning effectiveness held constant. GIFT similarly lends itself to cost and effectiveness analyses of other modules including authoring tools within and external to GIFT.

In sum, both GIFT and cost analyses are applicable to ITS development and design. GIFT in particular concerns all aspects of these systems – including cost analyses focused on assessment to advise decision

makers concerned with design, development, implementation, and assessment of technology-based tutoring. Cost analysis should become a routine, essential component in assessing return on investments in the adoption and application of GIFT products.

This chapter provides an introductory overview of cost-analysis as it might be applied to GIFT and ITSs. Additional, more comprehensive resources are available from Phillips (2011) who provides further guidance for cost analysis in training and Levin and McEwan (2001) who do so for education. Harris (2009) is also recommended for its discussion of effect sizes combined with cost analysis.

References

Carter, G. & Trollip, S. (1980). A constrained maximization extension to incremental transfer effectiveness, or, how to mix your training technologies. *Human Factors*, *22(2)*, 141–52.

Cohn, J. & Fletcher, J.D. (2010). What is a Pound of Training Worth? *Proceedings of the InterService/Industry Training, Simulation and Education Annual Conference*. Arlington, VA: National Training and Simulation Association.

Fletcher, J. D. (2009). Education and Training Technology in the Military. *Science, 323*, 72-75.

Fletcher, J. D. & Chatham, R. E. (2010). Measuring Return on Investment in Military Training and Human Performance. In, P. E. O'Connor & J V. Cohn (Eds.) *Human Performance Enhancements In High-Risk Environments* (pp. 106-128). Santa Barbara, CA: Praeger,/ABC-CLIO.

Fletcher, J.D., Hawley, D.E. & Piele, P.K. (1990). Costs, effects, and utility of microcomputer assisted instruction in the classroom. *American Educational Research Journal*, *27*, 783-806.

Harris, D.N. (2009). Toward Policy-Relevant Benchmarks for Interpreting Effect Sizes: Combining Effects with Costs. *Educational Evaluation and Policy Analysis, 31*, 3-29.

Holman, G. J. (1979). *Training Effectiveness of the CH-47 Flight Simulator* (ARI-RR-1209). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences. (ADA 072 317)

Hummel-Rossi, B. & Ashdown, J. (2002). The state of cost-benefit and cost effectiveness analyses in education. *Review of Educational Research*, *72*, 1-30.

Jones, M.B. & Kennedy, R.S. (1996). Isoperformance curves in applied psychology. *Human Factors*, 38 (1), 167–82.

Kearsley, G. (1982). *Costs, Benefits & Productivity in Training Systems*. Reading, MA: Addison-Wesley.

Knapp, M.I. & Orlansky, J. (1983). *A Cost Element Structure for Defense Training* (IDA Paper P-1709). Institute for Defense Analyses, Alexandria, Virginia (ADA 139 164).

Levin, H.A. (1983). *Cost-Effectiveness: A Primer*. Sage Publications, Beverly Hills, California.

Levin, H.M. & McEwan, P.J. (2001) *Cost-Effectiveness Analysis*. Thousand Oaks, CA: Sage.

Mishan, E.J. & Quah, (E.) (2007). *Cost Benefit Analysis*. London: Routledge.

Morrison, J.E. & Holding, D.H. (1990). *Designing a Gunnery Training Strategy (Technical Report 899)*. Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (ADA 226 129)

Phillips, J.J. (2011). *Return on Investment in Training and Performance Improvement Programs, (2nd Edition)*. Oxford, UK: Butterworth-Heinemann.

Rice, J.K. (1997). Cost analysis in education: Paradox and possibility. *Educational Evaluation and Policy Analysis*, *19*, 309-317.

Ross, J.A., Barkaoui, K. & Scott, G. (2007). Evaluations that consider the cost of educational programs: The contribution of high quality studies. *American Journal of Evaluation*, *28*, 477-492.

Sassone, P.G. & Schaffer, W.A. (1978). *Cost-Benefit Analysis: A Handbook*. New York: Academic Press.

Shannon, R.E. 1975. *Systems Simulation: The Art and the Science*. Englewood Cliffs, NJ: Prentice-Hall.

Sottilare, R.A., Brawner, K.W., Goldberg, B.S. & Holden, H.K. (2012). *The Generalized Intelligent Framework for Tutoring (GIFT)*. Orlando, FL: U.S. Army Research Laboratory – Human Research & Engineering Directorate (ARL-HRED).

Taylor, H.L., Talleur, D.A., Emanuel, T.W.,Jr., Rantanen, E.M., Bradshaw,G.L. & Phillips, S.I. (2002). *Incremental Training Effectiveness of Personal Computers Used for Instrument Training* (Final Technical Report ARL-02-5/NASA-02-3). Savoy, IL: University of Illinois, Aviation Research Lab Institute of Aviation..

Thompson, M.S. (1980). *Benefit-Cost Analysis for Program Evaluation*. Thousand Oaks, CA: Sage.

Towne, D.M. (1995). *Learning and Instruction in Simulation Environments*. Englewood Cliffs, NJ: Educational Technology Publications.

United States Department of Defense (1992). *DoD 5000.4-M, Cost Analysis Guidance and Procedures*. Washington, D.C.: Department of Defense

BIOGRAPHIES

Editors

Benjamin Goldberg
Benjamin Goldberg, Ph.D. is a member of the Learning in Intelligent Tutoring Environments (LITE) Lab at the U.S. Army Research Laboratory's (ARL), Human Research and Engineering Directorate (HRED), Simulation and Training Technology Center (STTC) in Orlando, FL. He has been conducting research in the Modeling & Simulation community for the past five years with a focus on adaptive learning and how to leverage Artificial Intelligence tools and methods for adaptive computer-based instruction. Currently, he is the LITE Lab's lead scientist on instructional strategy research within adaptive training environments. Dr. Goldberg is a Ph.D. graduate from the University of Central Florida in the program of Modeling & Simulation. Prior to employment with ARL, he held a Graduate Research Assistant position for two years in the Applied Cognition and Training in Immersive Virtual Environments (ACTIVE) Lab at the Institute for Simulation and Training. Dr. Goldberg's work has been published across several well-known conferences, with recent contributions to the Human Factors and Ergonomics Society (HFES), Artificial Intelligence in Education and Intelligent Tutoring Systems (ITS) proceedings, and to the *Journal of Cognitive Technology*.

Art Graesser
Dr. Art Graesser is a professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis and is an Honorary Research Fellow at the University of Oxford. His primary research interests are in cognitive science, discourse processing, and the learning sciences. More specific interests include knowledge representation, question asking and answering, tutoring, text comprehension, inference generation, conversation, reading, memory, emotions, computational linguistics, artificial intelligence, human-computer interaction, learning technologies with animated conversational agents (such as AutoTutor and Operation ARA), and automated analyses of texts at multiple levels (such as Coh-Metrix, and Question Understanding AID -- QUAID). He served as editor of the journal *Discourse Processes* (1996–2005) and *Journal of Educational Psychology* (2009-2014). His service in professional societies includes president of the Empirical Studies of Literature, Art, and Media (1989-1992), the Society for Text and Discourse (2007-2010), the International Society for Artificial Intelligence in Education (2007-2009), and the Federation of Associations for Behavioral and Brain Sciences Foundation (2012-13). In addition to receiving major lifetime research achievement awards from the Society for Text and Discourse and University of Memphis, he received an award in 2011 from the American Psychological Association on Distinguished Contributions of Applications of Psychology to Education and Training.

Xiangen Hu
Dr. Xiangen Hu is Dunavant professor in the Department of Psychology and Department of Electronic and Computer Engineering at The University of Memphis (UofM), senior researcher at the Institute for Intelligent Systems (IIS) at the UofM, and visiting professor at Central China Normal University (CCNU). Dr. Hu received his MS in Applied Mathematics from Huazhong University of Science and Technology, MA in Social Sciences, and Ph.D. in Cognitive Sciences from the University of California, Irvine. Currently, Dr. Hu is the director of the Cognitive Psychology program at the UofM, the Director of Advanced Distributed Learning (ADL) Center for Intelligent Tutoring Systems (ITS) Research &

Development, and the senior researcher in the Chinese Ministry of Education's Key Laboratory of Adolescent Cyberpsychology and Behavior.

Dr. Hu's primary research areas include Mathematical Psychology, Research Design and Statistics, and Cognitive Psychology. More specific research interests include General Processing Tree (GPT) models, categorical data analysis, knowledge representation, computerized tutoring, and advanced distributed learning. Dr. Hu receives funding for the above research from the US National Science Foundation (NSF), US Institute for Education Sciences (IES), ADL of the US Department of Defense (DoD), US Army Medical Research Acquisition Activity (USAMRAA), US Army Research Laboratories (ARL), US Office of Naval Research (ONR), UofM, and CCNU.

Robert Sottilare

Dr. Robert A. Sottilare serves as the Chief Scientist at the SFC Paul Ray Smith Simulation & Training Technology Center (STTC) within the Army Research Laboratory's Human Research and Engineering Directorate (ARL-HRED). He also leads adaptive tutoring research within ARL's Learning in Intelligent Tutoring Environments (LITE) Laboratory where the focus of his research is in automated authoring, instructional management, and analysis tools and methods for intelligent tutoring systems. His work is widely published and includes articles in the *Cognitive Technology Journal* and the *Educational Technology Journal*. Dr. Sottilare is a co-creator of the *Generalized Intelligent Framework for Tutoring* (GIFT) and the chief editor for the Design Recommendations for ITSs book series. He received his doctorate in Modeling & Simulation from the University of Central Florida (UCF) with a focus in intelligent systems. He is a visiting lecturer at both UCF and the United Stated Military Academy. In 2012, Dr. Sottilare was honored as the inaugural recipient of the U.S. Army Research Development & Engineering Command's *Modeling & Simulation Lifetime Achievement Award*.

Authors

Vincent Aleven

Dr. Vincent Aleven is an Associate Professor in Carnegie Mellon's Human-Computer Interaction Institute, and has 20 years of experience in research and development of advanced learning technologies based on cognitive science theory. His research focuses on metacognition, authoring tools, and the use of tutoring technology in ill-defined domains. Dr. Aleven and colleagues created CTAT, a suite of efficient, easy-to-learn, and easy-to-use authoring tools for intelligent tutoring systems, including a new paradigm called "example-tracing tutors." CTAT is living proof that end-user programming techniques can dramatically increase the cost- effectiveness of tutor authoring even for non-programmers. Dr. Aleven is a member of the Executive Committee of the Pittsburgh Science of Learning Center (PSLC), an NSF-sponsored research center spanning Carnegie Mellon and the University of Pittsburgh. He is a co-founder of Carnegie Learning, Inc., a Pittsburgh-based company that markets Cognitive Tutor™ math courses. He was the Program Committee Co-Chair of the 2010 International Conference on Intelligent Tutoring Systems. He is Co-Editor in Chief of the *International Journal of Artificial Intelligence in Education.* He has been or is PI on 7 major research grants and co-PI on 8 others. He has authored over 140 peer-reviewed publications.

Ivon Arroyo

Ivon Arroyo is Assistant Professor in Learning Sciences and Technologies at Worcester Polytechnic Institute, Worcester, MA. She has designed and implemented tutoring systems for mathematics education, and carried out research on how students learn and perceive mathematics with intelligent tutors at the K-12 level in public school settings. She is PI or Co-PI of NSF and U.S. DoEd grants that predicted student emotions in real time and designed mathematics tutors for students with learning disabilities, and for girls. She holds an Ed.D. in mathematics and science education and M.S. and B.S. degrees in computer science.

She is the author of over 100 research articles, is a Fulbright Fellow, was an elected member of the executive committee of AIED, and received several best paper awards (AIED 2009, 2010 and EDM 2010), one of them entitled "Emotion Sensors Go to School."

Whitney Baer
Whitney Baer is a staff member at the Center for the Study of Adult Literacy (CSAL). She has a Bachelor of Arts in English and History from Colgate University. Prior to joining the team at the University of Memphis, she worked for seven years in educational publishing. In that capacity, she gained experience with course redesign and learning management systems; as well as educational research and development. Whitney coordinates the collaborative efforts of the CSAL team at the Institute for Intelligent Systems at The University of Memphis. She is responsible for helping to develop curriculum and managing the process of interpreting scripted conversation into the AutoTutor framework.

Ryan Baker
Ryan Shaun Baker is Associate Professor of Cognitive Studies at Teachers College, Columbia University. He earned his Ph.D. in Human-Computer Interaction from Carnegie Mellon University. Baker was previously Assistant Professor of Psychology and the Learning Sciences at Worcester Polytechnic institute, and he served as the first Technical Director of the Pittsburgh Science of Learning Center DataShop, the largest public repository for data on the interaction between learners and educational software.

He is currently serving as the founding President of the International Educational Data Mining Society, and as Associate Editor of the *Journal of Educational Data Mining*. His research combines educational data mining and quantitative field observation methods in order to better understand how students respond to educational software, and how these responses impact their learning. He studies these issues within intelligent tutors, simulations, multi-user virtual environments, and educational games, within populations from pre-schoolers, to middle school students, to military trainees.

Gautam Biswas
Gautam Biswas is a Professor of Computer Science, Computer Engineering, and Engineering Management in the EECS Department and a Senior Research Scientist at the Institute for Software Integrated Systems at Vanderbilt University. He conducts research in Intelligent Systems with primary interests in analysis of complex embedded systems, and simulation-based environments for learning and instruction. This includes the Teachable Agents project, where students learn science by building causal models of natural processes and CTSiM, which exploits the synergy between computational thinking ideas and STEM learning to develop systems that help students learn science and math concepts by building simulation models. He has developed innovative educational data mining techniques for studying students' learning behaviors and linking them to metacognitive strategies. His research has been supported by funding from NASA, NSF, DARPA, and the US Department of Education. He has published extensively, and has over 400 refereed publications. Dr. Biswas is an associate editor of the *IEEE Transactions on Systems, Man, and Cybernetics*, *IEEE Transactions on Learning Technologies*, and the *Metacognition and Learning* journal. He is a Fellow of the IEEE, and member of the ACM, AAAI, and the Sigma Xi Research Society.

Nathaniel Blanchard
Nathaniel Blanchard is a research assistant in the Computer Science and Engineering department at the University of Notre Dame. He received his undergraduate degree in computer science from Hanover College. His research interests include detecting and responding to emotions during learning and the use of intelligent tutoring systems to facilitate learning. Currently, he is focusing on using machine learning to automatically detect mind wandering in learning environments using skin conductance and skin temperature signals. He is also examining the applicability of speech recognition software and acoustic features to

automatically categorize learning in classroom environments. Previous work has included designing and building a number of research projects including an aerobic exercise game at the University of North Carolina at Charlotte, a game environment for teaching introductory computer science topics at Hanover College, and a software library aimed at facilitating the writing of branching dialog for integration into video games.

Keith Brawner

Dr. Keith Brawner is a researcher for the Learning in Intelligent Tutoring Environments (LITE) Lab within the U. S. Army Research Laboratory's Human Research & Engineering Directorate (ARL-HRED). He has 8 years of experience within U.S. Army and Navy acquisition, development, and research agencies. He holds a Masters and PhD degree in Computer Engineering with a focus on Intelligent Systems and Machine Learning from the University of Central Florida. The foci of his current research is in machine learning, active and semi-supervised learning, realtime datastream processing, affective computing, adaptive training, and semi/fully automated user tools for adaptive training content.

Winslow Burleson

Winslow Burleson is an Associate Professor at New York University. He is the Founding Director of the Motivational Environments research group and author of over 100 scientific publications (including the "best paper" at AI in Ed 2009 and the 2011 UMUAI James Chen Award) and has been awarded 10 patents. In 2013, he was honored with a Google Faculty Research Award and in 2009 the National Academy of Engineering recognized him as, "One of the nation's brightest young engineering researchers and educators." He received his PhD from the MIT Media Lab, is a Kavli Fellow and serves on National Academy of Engineering, National Academies of Sciences, and National Science Foundation committees. Burleson has a BA in Biophysics from Rice University and an MSE in Mechanical Engineering from Stanford University.

Whitney Cade

Whitney Cade is an Experimental Psychology doctoral student at the University of Memphis with a concentration in Cognitive Psychology. She graduated from Rhodes College with a BA in Psychology. Working primarily in the Institute for Intelligent Systems, her primary interests include cognitive science and learning. Her research specifically focuses on expert human tutoring, large-grain pedagogical strategies, intelligent tutoring systems, pedagogical agents, image display techniques that support learning, and the application of machine learning in educational software.

Zhiqiang Cai

Mr. Zhiqiang Cai is a Research Assistant Professor with the Institute for Intelligent Systems at the University of Memphis. He has a Master of Science degree in computational mathematics received in 1985 from Huazhong University of Science and Technology, P. R. China. His current research interests are in algorithm design and software development for tutoring systems and natural language processing.

Mark Conley

Mark W. Conley is a professor of literacy at the University of Memphis. He specializes in human and technology-based tutoring and assessment in literacy, mathematics and science. He designed the Memphis Literacy Corps, the largest literacy tutoring effort implemented in the United States. Currently, he is designing tutoring programs for adult learners in community-based tutoring programs, including a program for potentially incarcerated youth. He has published numerous articles about literacy, disciplinary literacy, teacher education and assessment in journals like the *Harvard Educational Review*, *Theory Into Practice* and the *Journal of Adolescent and Adult Literacy*. He has published several books about disciplinary literacy, curriculum standards and assessment. He has also advised officials at state, national and international levels on curriculum standards and assessment. He holds a Certified Flight Instructor

certificate with an instrument rating and teaches and practices guitar building as part of his passion for teaching and learning.

Janice Connor
Janice Connor has a BA in Corporate Communications from Arcadia University, Glenside, PA and most recently worked as a Research Assistant for the Learning in Intelligent Tutoring Environments (LITE) Laboratory at the US Army Research Laboratory, Simulation and Training Technology Center in Orlando, FL. Prior, she worked for non-profit and educational organizations while also supporting her spouse's 24-year career as an Officer in the US Army. This is her first official publication and she is grateful for the opportunity.

Mark G. Core
Mark G. Core is a research scientist at the Institute for Creative Technologies (ICT) at the University of Southern California. He specializes in artificial intelligence in education working in ill-defined domains such as negotiation, cultural awareness and leadership. He received his Ph.D. from the University of Rochester in 2000 under the direction of Prof. Lenhart Schubert, and was a research fellow at the University of Edinburgh working with Prof. Johanna Moore until joining ICT in 2004. He worked in the area of computational linguistics specifically natural language understanding and discourse analysis. At the University of Edinburgh, Mark undertook analysis of successful human tutoring dialogues and while at ICT he has focused on analysis of learner writing. At ICT, he is also working on an evolving tutoring architecture incorporating technologies such as expert modeling, open learner modeling, experience manipulation, explainable artificial intelligence, natural language understanding, and natural language generation. Recent areas of research include authoring tools for tutoring systems, and physician training including interview and diagnosis skills.

Jeanine DeFalco
Jeanine A. DeFalco is a Doctoral Research Fellow in Cognitive Studies with a focus in Intelligent Technologies at Teachers College, Columbia University, working under a grant from the Army Research Lab. Jeanine's research interests include embodied cognition and the theory of cognitive extension as it relates to whole-task, game-based learning platforms for the development of skill and reasoning expertise. A member of Kappa Delta Pi, the international honors society for education, Jeanine has a Masters in Educational Theatre, Colleges and Communities, from the Steinhardt School, New York University, and a Masters in Drama Studies from The Johns Hopkins University.

Sidney D'Mello
Sidney D'Mello is an Assistant Professor in the departments of Computer Science and Psychology at the University of Notre Dame. His primary research interests are in the affective, cognitive, and learning sciences. More specific interests include affective computing, artificial intelligence in education, human-computer interaction, natural language understanding, and computational models of human cognition. He has co-edited five books and has published over 150 journal papers, book chapters, and conference proceedings in these areas. D'Mello's work on intelligent learning technologies including Affective AutoTutor, GazeTutor, ConfusionTutor, and GuruTutor has received seven outstanding paper awards at international conferences and has been featured in several media outlets including the *Wall Street Journal*. D'Mello serves on the executive board of the *International Artificial Intelligence in Education Society*, is a senior reviewer for the *Journal of Educational Psychology*, is an associate editor for *IEEE Transactions on Affective Computing*, and for *IEEE Transactions on Learning Technologies*. D'Mello received his PhD. in Computer Science from the University of Memphis in 2009. He also holds a M.S. in Mathematical Sciences and a B.S. in Electrical Engineering.

Paula Durlach

Paula J. Durlach received her Ph.D. in experimental psychology from Yale University in 1982, and subsequently held fellowship positions at the University of Pennsylvania and the University of Cambridge. From 1987 to 1994, she was an assistant professor of psychology at McMaster University and then went on to lead the exploratory consumer science team at Unilever Research Colworth Laboratory in the U. K. She returned to the U. S. in 2001 to join the U. S. Army Research Institute for the Behavioral and Social Sciences. Since April 2012, she has been the Deputy Director of the Advanced Distributed Learning Initiative. Dr. Durlach has received recognition for her work in experimental psychology and cognitive science at the Army Science Conference and from the Department of Army Research and Development. She is a Fellow of the Association for Psychological Science, and member of the Experimental Psychology Society, the Psychonomic Society, and the Society for Artificial Intelligence in Education. With Dr. Alan Lesgold, she co-edited the book, *Adaptive Technologies for Training and Education*, published in 2012, and has also published research in journals such as *International Journal of Artificial Intelligence in Education*, *Military Psychology*, *Computers in Human Behavior*, and *Human-Computer Interaction*.

Stephen E. Fancsali

Stephen E. Fancsali is a Research Scientist at Carnegie Learning, Inc. His work centers on topics in educational data mining and intelligent tutoring systems research, including student modeling, personalization and adaptation of educational software based on non-cognitive factors, and statistical and causal modeling of learner behavior, affect, and other factors from observational data, especially complex, "raw" log data from sources like intelligent tutoring systems. He also supports data-intensive, mathematics education research that uses Carnegie Learning's Cognitive Tutor as a platform for experimentation. Fancsali earned the Ph.D. in Logic, Computation, and Methodology from the Department of Philosophy at Carnegie Mellon University in 2013. His dissertation focused on theoretical and applied problems of variable construction (i.e., feature extraction) from fine-grained, raw data in support of causal inference and discovery.

Shi Feng

Shi Feng is a PhD candidate at the University of Memphis. She successfully defended her Master's at the University of Memphis under the supervision of Dr. Art Graesser. She joined the Center for the Study of Adult Literacy in 2012 for developing AutoTutor modeled framework for helping struggling adult readers. Her other projects include disengagement during reading, and mind wandering during discourse comprehension. Her current interest includes computational linguistics, discourse processing and comprehension, text inferences, engagement during reading, and developing interesting texts for educational learning.

Dexter Fletcher

J. D. Fletcher is a member of the senior research staff at the Institute for Defense Analyses where he specializes in personnel and human performance issues. His graduate degrees are in computer science and educational psychology from Stanford University where, as a research associate, he directed projects for the Institute for Mathematical Studies in the Social Sciences. He has held university positions in psychology, computer science, and systems engineering and government positions in Navy and Army Service Laboratories, the Defense Advanced Research Projects Agency, and the White House Office of Science and Technology Policy. He has served on science and technology advisory panels for the Defense Science Board, Army Science Board, Naval Studies Board, Air Force Scientific Advisory Board, National Science Foundation, National Academy of Sciences, and the National Academy of Engineering. He has designed computer-based instruction programs used in public schools and training devices used in military training. He is a Fellow of the American Educational Research Association and three divisions of the American Psychological Association. His research interests include intelligent tutoring systems, synthetic environ-

ments in education and training, mobile performance aids, analyses of skilled behavior, and cost-effectiveness analyses of education and training.

Michael Hoffman
Michael Hoffman is a software engineer at Dignitas Technologies with over eight years of experience in software development. Upon graduating from the University of Central Florida with a B.S. in Computer Science, he spent a majority of his time on various OneSAF development activities for SAIC. He worked on the DARPA Urban Challenge, where he provided a training environment for the robot by simulating AI traffic and the various sensors located on Georgia Tech's Porsche Cayenne in a OneSAF environment. Soon after earning a Master of Science degree from the University of Central Florida, Michael found himself working at Dignitas. Both at Dignitas and on his own time, Michael has created several iPhone applications. One application, called the Tactical Terrain Analysis app, provides mobile situation awareness and can be used as a training tool for various real world scenarios. More recently he has worked to determine if unobtrusive sensors can be used to detect an individual's mood during a series of computer interactions. Michael excels in integrating both software and hardware systems such as third party simulations and sensors. Michael has been the lead software engineer on GIFT since its inception in 2011.

Heather Holden
Dr. Holden is currently a researcher in the Learning in Intelligent Tutoring Environments (LITE) Lab within the U.S. Army Research Laboratory – Human Research and Engineering Directorate (ARL-HRED) – Simulation and Training Technology Center (STTC) in Orlando, Florida. The focus of her research is in learner modeling, artificial intelligence, and computer-based tutoring system application to education and training. Her research interests also include technology acceptance and Human-Computer Interaction. Dr. Holden previously served as an Information Technology Specialist for the Social Security Administration (SSA) National Computing Center in Woodlawn, Maryland. Dr. Holden earned her Doctorate and Masters in Information Systems from the University of Maryland, Baltimore County. She also has a graduate certificate in Instructional Technology from the same university. Her doctoral research evaluated the relationship between teachers' technology acceptance and usage behaviors to better understand the perceived usability and utilization of job-related technologies. Her work has been published in the *Journal of Research on Technology in Education*, the *International Journal of Mobile Learning and Organization*, the *Interactive Technology and Smart Education Journal*, and several relevant conference proceedings. Her doctoral work has been continued by other researchers in academia. Dr. Holden also possesses a BS in Computer Science from the University of Maryland, Eastern Shore.

John S. Kinnebrew
John S. Kinnebrew is a Research Scientist at the Institute for Software Integrated Systems at Vanderbilt University. He received his B.A. in Computer Science from Harvard University and his Ph.D. in Computer Science from Vanderbilt University. His research interests include data mining, user modeling, and coordination in multi-agent systems. He is currently involved in a variety of computer-based learning environment projects, where his research focuses on data mining and machine learning for modeling human learning behaviors, including metacognition and self-regulated learning strategies.

Ken Koedinger
Dr. Kenneth Koedinger is Professor of Human-Computer Interaction and Psychology at Carnegie Mellon. His research has contributed new principles and techniques for the design of educational software and has produced basic cognitive science research results on the nature of student thinking and learning. Dr. Koedinger is a co-founder of Carnegie Learning (http://carnegielearning.com) and the CMU Director of LearnLab (http://learnlab.org). LearnLab is supporting Big Data investigations in education and, more generally, leverages cognitive and computational approaches to support researchers in investigating the instructional conditions that cause robust student learning. See http://pact.cs.cmu.edu/koedinger.html for more information.

Susanne Lajoie

Dr. Susanne Lajoie is a Professor and Canadian Research Chair Tier 1 in Advanced Technologies for Learning in Authentic Settings in the Department of Educational and Counselling Psychology at McGill University and a member of the Centre for Medical Education. She is a Fellow of the American Psychological Association, appointed for her outstanding contributions to the field of Psychology as well as a Fellow of the American Educational Research Association. She received her Doctorate from Stanford University in 1986. Dr. Lajoie is a recipient of the McGill Carrie Derick Award for graduate supervision and teaching. Dr. Lajoie is the Director of the Learning Environments Across Disciplines partnership grant funded by the Social Sciences and Humanities Research Counsel in Canada. Her research involves the design of technology rich learning environments for educational and professional practices. She explores how theories of learning and affect can be used to guide the design of advanced technology rich learning environments in different domains, i.e. medicine, mathematics, history, etc. These environments serve as research platforms to study student engagement and problem solving in authentic settings. She uses a cognitive approach to identify learning trajectories that help novice learners become more skilled in specific areas and designs computer tools to enhance self-regulation, memory, and domain-specific learning. She has numerous publications and has been invited to present her research worldwide including Australia, France, Germany, Hong Kong, Korea, Singapore, Spain, Sweden, Taiwan, Mexico, the UK and the Ukraine.

H. Chad Lane

H. Chad Lane is a Research Scientist and Director for Learning Sciences Research at the USC Institute for Creative Technologies (ICT). His work focuses on the application of artificial intelligence and entertainment technologies to educational problems. He has published over 40 papers in areas including educational games, pedagogical agents, scaffolding/feedback, and virtual environments for learning. Chad received his PhD in Computer Science in 2004 from the University of Pittsburgh, and MS in the Computer Sciences from the University of Wisconsin-Madison prior to that. Recently, Chad was the Program Co-Chair for the 16th International Conference on Artificial Intelligence in Education (AIED). He also serves on the executive committee of the AIED Society (elected position), as an associate editor for several major educational technology journals, and as an advisor for the NSF Cyberlearning CIRCL center. More information is available on his website: http://people.ict.usc.edu/~lane

James Lester

James C. Lester is Distinguished Professor of Computer Science at North Carolina State University. His research focuses on transforming education with technology-rich learning environments. Utilizing artificial intelligence, game technologies, and computational linguistics, he designs, develops, fields, and evaluates next-generation learning technologies for K-12 science, literacy, and computer science education. His work on personalized learning ranges from game-based learning environments and intelligent tutoring systems to affective computing, computational models of narrative, and natural language tutorial dialogue. He received his B.A. (Highest Honors, Phi Beta Kappa), M.S.C.S., and Ph.D. in computer science from the University of Texas at Austin. He received his B.A. in history from Baylor University. He has served as Program Chair for the International Conference on Intelligent Tutoring Systems, the International Conference on Intelligent User Interfaces, and the International Conference on Foundations of Digital Games, on the editorial board of Metacognition and Learning, and as Editor-in-Chief of the *International Journal of Artificial Intelligence in Education*. He has been recognized with a National Science Foundation CAREER Award and several Best Paper Awards.

Eleni Lobene

Eleni Lobene is a Research Psychologist in the Department of Computer Science at North Carolina State University. She received her Ph.D. in Industrial/Organizational Psychology from North Carolina State University in 2011, where her research focused on K-12 teacher motivations and perceptions. Prior to joining the Department of Computer Science, she served as a Research Assistant at the Friday Institute for

Educational Innovation, focusing on the assessment and evaluation of game-based learning environments for middle school education. She has served as the Managing Editor for the *International Journal of Artificial Intelligence*, an instructor for undergraduate courses in the Department of Psychology at North Carolina State University, and a consultant to local and global organizations focusing on improving organizational efficiency and implementing data-based change. Dr. Lobene's research interests span advanced learning technologies, assessment, and career development.

Yanjin Long
Yanjin Long is a Ph.D. student in the Human-Computer Interaction Institute at Carnegie Mellon University (CMU), focusing on learning sciences and learning technologies. She is also an associate of the Program in Interdisciplinary Education Research (PIER) at CMU. Yanjin received her B.S. degree in Psychology from Beijing Normal University (2008) and M.A. of Cognitive Studies in Education from Teachers College, Columbia University (2010). Her current research focuses on two main lines: designing and developing online tools to foster students' Self-Regulated Learning; and conducting classroom experiments to empirically evaluate the effectiveness of these new technologies. In her research, she has shown that support for self-assessment (either in the form of a skill-diary or an Open Learner Model with self-assessment support) can help students achieve better learning results. Her work on skill diaries won her the Conference Best Student Paper Award during the 16th International Conference on Artificial Intelligence in Education, AIED 2013, held in Memphis, TN, July 2013.

Chip Morrison
Dr. Chip Morrison is a Research Assistant Professor at IIS. A graduate of Dartmouth College, Dr. Morrison holds an M.A. in Language Studies from the University of Hong Kong and an Ed.D. in Human Development from the Harvard Graduate School of Education. Dr. Morrison has spent his entire professional career in education, including ten years teaching English as a Second Language in Hong Kong, several years as an educational software developer, and more than 15 years leading research and development efforts connected with the comprehensive school reform movement in the United States. As a Senior Scientist at Bolt, Beranek and Newman, Dr. Morrison helped found Co-nect, a comprehensive school reform model. Among other contributions, he established and directed the Co-nect Critical Friends Program, a review process used by hundreds of schools nationwide. In 2006, he joined Dr. Ron Ferguson's Tripod Project, part of the Achievement Gap Initiative at Harvard's Kennedy School of Government, and subsequently spent a year as lead School Quality Reviewer for the New York City Public Schools. Since coming to the University of Memphis in 2008, he has helped bring in grants and contracts worth more than $500K annually, including a five-year, $3.5M grant from the U.S. Department of Education to evaluate a large-scale science education initiative run by The Smithsonian Institution. He is currently Co-Principal Investigator and IIS Project Director for a contract funded by the Advanced Distributed Learning (ADL) Initiative (U.S. Department of Defense) involving the analysis of some 250,000 human-human tutorial dialog transcripts accumulated by Tutor.com, a leading provider of online tutorial services for children and young adults.

Bradford Mott
Bradford Mott is a Senior Research Scientist in the Department of Computer Science at North Carolina State University. Prior to joining North Carolina State University, he served as Technical Director at Emergent Game Technologies where he created cross-platform middleware solutions for video game development, including solutions for the PlayStation 3, Wii, and Xbox 360. Dr. Mott received his Ph.D. in Computer Science from North Carolina State University in 2006, where his research focused on intelligent game-based learning environments. His current research interests include computer games, computational models of interactive narrative, and intelligent game-based learning environments.

Kasia Muldner

Kasia Muldner received her Ph.D. from the Department of Computer Science at the University of British Columbia, where she designed and evaluated a computational tutor that supported students during analogical problem solving. She is a Research Scientist in the Department of Computing, Informatics, and Decision Systems Engineering at Arizona State University. Her work falls into the intersection of Human-Computer-Interaction and Artificial Intelligence, dealing with the design and evaluation of interactive educational technologies that aim to help students learn effectively though personalized support. She is particularly interested in technologies that support high level student states related to meta-cognition, affect, and creativity.

Benjamin Nye

Benjamin D. Nye is a research assistant professor at the University of Memphis at the Institute for Intelligent Systems. His current focus is on intelligent tutoring system (ITS) architectures, with a focus on lowering barriers to developing and adopting ITS technology. His primary research project is the Office of Naval Research (ONR) STEM Grand Challenge, where he is researching natural language tutoring modules called Sharable Knowledge Objects (SKO). Ben is also involved in cognitive agent-based architectures. Ben's thesis topic was "Modeling Memes: A Memetic View of Affordance Learning," which examined memes theoretically and computationally through a model that synthesized Shannon Information Theory and Observational Learning from Bandura's Socio-Cognitive Learning Theory.

Jaclyn Ocumpaugh

Jaclyn Ocumpaugh (jo2424@columbia.edu) is a Research Associate specializing in the Learning Sciences at Teachers College, Columbia, where her research focuses on improving the affective support provided by educational software through a combination of field work and learning analytics techniques. Her PhD (Sociolinguistics, Michigan State University), examined ethnographic patterns within an ethnic minority group, correlating them with changes in the acoustic patterns of the vowel system. In her new field, she seeks integrate these same skills into more common practices of the learning sciences, improving the representation and understanding of how cultural factors impact learning. She recently completed Post-doctoral Fellowship in Learning Sciences at Worcester Polytechnic Institute.

Andrew Olney

Andrew Olney is presently an Associate Professor in the Institute for Intelligent Systems / Department of Psychology at the University of Memphis and Director of the Institute for Intelligent Systems. Dr. Olney received a B.A. in Linguistics with Cognitive Science from University College London in 1998, an M.S. in Evolutionary and Adaptive Systems from the University of Sussex in 2001, and a Ph.D. in Computer Science from the University of Memphis in 2006. Dr. Olney's primary research interests are in natural language interfaces. Specific interests include vector space models, dialogue systems, unsupervised grammar induction, robotics, and intelligent tutoring systems.

Dr. Olney frequently serves as program committee member and journal reviewer in the fields of cognitive science, artificial intelligence, and education. Together with his collaborators, Dr. Olney has been awarded $9.3 million from federal funding agencies including the National Science Foundation, the Institute for Education Sciences, and the Department of Defense. His research has been featured in WIRED Magazine, the New York Times, the *Wall Street Journal*, the Discovery Science Channel, and BBC Radio 4. Dr. Olney was awarded first place in an international robotics competition for the PKD Android (AAAI, 2005) and received the Early Career Research Award from the University of Memphis.

Philip Pavlik

Philip I. Pavlik Jr. is currently an assistant professor of Psychology at the University of Memphis Institute for Intelligent Systems. Dr. Pavlik received a BA from the University of Michigan in Economics and a PhD from Carnegie Mellon University where he studied Cognitive Psychology with John Anderson

(developer of the ACT-R cognitive modeling system) and received a Neuroscience certificate from the Center for the Neural Basis of Cognition. With Anderson, Pavlik has pioneered changes in the ACT-R theory that have allowed his research to use this theory to quantitatively optimize the learning of information for tasks such as flashcard learning. From this foundation, his work with Ken Koedinger has developed to focus on problem solving, schema learning, optimal transfer, effects of motivational constructs, and student strategy use. His methodologies include theory development, experimentation, mathematical modeling, and educational applications. Pavlik has received more than 1.5 million dollars in grant awards from the Institute for Educational Sciences, the National Science Foundation, and other sources.

Eric Poitras
Dr. Eric Poitras is a Postdoctoral Fellow of the Learning Environments Across Disciplines research partnership in the Department of Educational and Counselling Psychology at McGill University. He has recently been appointed as an Assistant Professor in the Instructional Design and Educational Technology program in the Department of Educational Psychology at the University of Utah. Dr. Poitras has received his MA (2010) and PhD (2013) in Educational Psychology from the Learning Sciences program from McGill University. Currently, Dr. Poitras is a member of the Advanced Technologies for Learning in Authentic Settings laboratory. His research interests include educational technology, social sciences education, self-regulated learning, inquiry-based learning, and educational data mining. In particular, the role of computer- and mobile-based applications in support of learning about history, and how to improve the adaptive capabilities of these systems. Dr. Poitras receives funding for this research through the Social Sciences and Humanities Research Council of Canada, the Learning Environments Across Disciplines research partnership, The History Education Network, and the Digital Humanities Engine.

Charles Ragusa
Charles Ragusa is a senior software engineer at Dignitas Technologies with over thirteen years of software development experience. After graduating from University of Central Florida with a B.S. in computer science, Mr. Ragusa spent several years at SAIC working on a variety of R&D projects in roles ranging from software engineer and technical/integration lead to project manager. Noteworthy projects include the 2006 DARPA Grand Challenge as an embedded engineer with the Carnegie Mellon Red Team, program manager of the SAIC CDT/MRAP IR&D project, and lead engineer for Psychosocial Performance Factors in Space Dwelling Groups. Since joining Dignitas Technologies in 2009, he has held technical leadership roles on multiple projects, including his current role as the principal investigator for the GIFT project.

Mark Riedl
Dr. Mark Riedl is an Associate Professor in the Georgia Tech School of Interactive Computing and director of the Entertainment Intelligence Lab. Dr. Riedl's research focuses on the intersection of artificial intelligence, storytelling, and virtual worlds. Dr. Riedl seeks to understand how computational systems can represent, reason about, and create narratives and interactive stories. His primary research is in automated narrative generation, the creation of fictional narratives by intelligent systems. He also explores how intelligent systems can improve human experiences in games and virtual worlds through dynamic game adaptation and automated game design. Dr. Riedl earned a PhD degree in 2004 from North Carolina State University, where he developed intelligent systems for generating stories and managing interactive user experiences in computer games. From 2004 to 2007, Dr. Riedl was a Research Scientist at the University of Southern California Institute for Creative Technologies where he researched and developed interactive, narrative-based training systems. Dr. Riedl joined the Georgia Tech College of Computing in 2007 and in 2011 he received a DARPA Young Faculty Award and NSF CAREER Award for his work on artificial intelligence, narrative, and virtual worlds.

Steve Ritter

Steven Ritter is Chief Scientist at Carnegie Learning. Dr. Ritter received his doctorate in cognitive psychology from the Carnegie Mellon University and worked with John Anderson and others to develop and evaluate the Intelligent Tutoring Systems that became the basis for Carnegie Learning's products. He was one of the co-founders of Carnegie Learning.

Dr. Ritter is the author of numerous papers on the design, architecture and evaluation of educational technology and served on the education board of the Software and Information Industry Association. His evaluation work has been recognized by the What Works Clearinghouse as fully satisfying their requirements for rigorous evaluation. In his role as Chief Scientist, Dr. Ritter directs all projects regarding research on the effectiveness of Cognitive Tutor products and guides development projects focused on improving the effectiveness of mathematics curricula. Dr. Ritter also serves as Chief Product Architect, setting the direction of future Cognitive Tutor products.

Ido Roll

Dr. Ido Roll is the Senior Manager for Research and Evaluation in the Centre for Teaching, Learning, and Technology at the University of British Columbia (UBC), and he is a researcher with the Pittsburgh Science of Learning Centre. Ido graduated from the Human-Computer Interaction Institute and the Program for Interdisciplinary Education Research in Carnegie Mellon University.

Ido studies how interactive learning environments support students in becoming more competent, curious, creative, and collaborative learners in classroom and online environments. His work focuses on cognitive and non-cognitive factors across different time scales, from minutes (in problem solving environments and simulations) to months (in MOOCs and learning management systems). His research utilizes a variety of methodologies from the fields of cognitive science, the learning sciences, artificial intelligence, learning analytics, education, and human-computer interaction. His publications in these fields have won numerous awards, and his research has been funded by the National Science Foundation (NSF), the Social Sciences and Humanities Research Council of Canada (SSHRC), the Gordon and Betty Moore Foundations (GBMF), and others.

Jonathan Rowe

Jonathan Rowe is a Research Scientist in the Department of Computer Science at North Carolina State University. He received Ph.D. and M.S. degrees in Computer Science from North Carolina State University. He received a B.S. degree in Computer Science from Lafayette College. His research is in the areas of artificial intelligence and human-computer interaction for advanced learning technologies, with an emphasis on game-based learning environments. He is particularly interested in intelligent tutoring systems, user modeling, educational data mining, and computational models of interactive narrative. Jonathan has led development efforts on several game-based learning projects, including Crystal Island: Lost Investigation, which was nominated for Best Serious Game at the 2012 Unity Awards and the 2012 I/ITSEC Serious Games Showcase and Challenge. His research has also been recognized with several best paper awards, including best paper at the Seventh International Artificial Intelligence and Interactive Digital Entertainment Conference and best paper at the Second International Conference on Intelligent Technologies for Interactive Entertainment.

Vasile Rus

Dr. Vasile Rus is an Associate Professor of Computer Science with a joint appointment in the Institute for Intelligent Systems (ITS) whose areas of expertise are computational linguistics, artificial intelligence, software engineering, and computer science in general. His research areas of interest include question answering and asking, dialogue-based intelligent tutoring systems (ITSs), knowledge representation and reasoning, information retrieval, and machine learning. For the past 10 years, Dr. Rus has been heavily involved in various dialogue-based ITS projects including systems that tutor students on science topics

(DeepTutor), reading strategies (iSTART), writing strategies (W-Pal), and metacognitive skills (MetaTutor). Currently, Dr. Rus leads the development of the first intelligent tutoring system based on learning progressions, DeepTutor (www.deeptutor.org). He has coedited two books, received several Best Paper Awards, and authored more than 90 publications in top, peer-reviewed international conferences and journals. He is currently Associate Editor of the *International Journal on Artificial Intelligence Tools*.

Wolfgang Schnotz

Dr. Wolfgang Schnotz is a Professor of General and Educational Psychology at University of Koblenz-Landau. His focus in teaching is on cognitive psychology and instructional psychology. He also teaches language and cognition as well as visualization with a focus on new media.

Dr. Schnotz received his PhD from the Technical University Berlin. He held positions at University of Tübingen, University of Bielefeld, University of Vienna, and University of Jena. He is now the head of the Department of General and Educational Psychology, the head of the Multimedia Research Centre and the head of the (German-Science-Foundation supported) Graduate School on Teaching and Learning Processes at the University of Koblenz-Landau.

Wolfgang Schnotz was Chief Editor of the international journal *Learning and Instruction*, member of the International Reading Expert Group for PISA 2009 and is editorial board member of numerous journals. He has published widely in the field of reading and listening comprehension, learning from text, comprehension of graphics, learning with hypermedia and learning from animation. He runs currently various research projects on text-picture-integration skills and coherence formation from conflicting information funded by the German Science Foundation.

James R. Segedy

James R. Segedy is a Ph.D. Candidate at the Institute for Software Integrated Systems at Vanderbilt University. He has an M.S. in Computer Science and has been teaching a college preparatory course to high school students since 2009. His research focuses on the design and implementation of technology-based tools for preparing students to productively engage in and manage difficult, long term, and open-ended problem solving tasks. This has involved conducting research into both (1) new techniques for understanding and characterizing learners' behaviors during open-ended problem solving, and (2) new approaches to designing effective, automated tutoring strategies for virtual tutors embedded within these environments.

Valerie Shute

Valerie Shute is the Mack & Effie Campbell Tyner Endowed Professor in Education in the Department of Educational Psychology and Learning Systems at Florida State University. Before coming to FSU in 2007, she was a principal research scientist at Educational Testing Service where she was involved with basic and applied research projects related to assessment, cognitive diagnosis, and learning from advanced instructional systems. Her general research interests hover around the design, development, and evaluation of advanced systems to support learning—particularly related to 21st century competencies. An example of current research involves using immersive games with stealth assessment to support learning—of cognitive and non-cognitive knowledge, skills, and dispositions. Her research has resulted in numerous grants, journal articles, books, chapters in edited books, a patent, and a couple of books (e.g., Shute & Ventura, 2013—*Measuring and supporting learning in games: Stealth assessment,* MIT Press.)

Anne Sinatra

Anne M. Sinatra, Ph.D. is an Oak Ridge Associated Universities Post Doctoral Fellow in the Learning in Intelligent Tutoring Environments (LITE) Lab at the U.S. Army Research Laboratory's (ARL) Simulation and Training Technology Center (STTC) in Orlando, FL. The focus of her research is in cognitive and human factors psychology. She has specific interest in how information relating to the self and about

those that one is familiar with can aid in memory, recall, and tutoring. Her dissertation research evaluated the impact of using degraded speech and a familiar story on attention/recall in a dichotic listening task. Her work has been published in the journal *Interaction Studies*, and in the conference proceedings of the Human Factors and Ergonomics Society. Prior to becoming a Post Doc, Dr. Sinatra was a Graduate Research Associate with UCF's Applied Cognition and Technology (ACAT) Lab, and taught a variety of undergraduate Psychology courses. Dr. Sinatra received her Ph.D. and M.A. in Applied Experimental and Human Factors Psychology, as well as her B.S. in Psychology from the University of Central Florida.

Matthew Small

Dr. Matthew Small is a Research Consultant and Educational Game Developer working with various educational research entities including the College of Education at The Florida State University. His research interests center on re-designing the template for modern educational game development through embedding mechanisms for learning and assessment into engaging video games that rely on non-linear gameplay mechanics. His work focuses on STEM areas of content knowledge that are traditionally considered difficult to teach and measure in "fun" digital environments. Dr. Small has collaboratively developed educational software including *Newton's Playground* and *Codecraft*, games that teach quantitative physics and computational thinking, respectively, and are the focus of numerous publications as well as ongoing research.

Randall Spain

Randall Spain is a research psychologist at the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). He received his Ph.D. in human factors psychology from Old Dominion University. His research at ARI has focused on how to accelerate learning and skill acquisition using intelligent learning environments. Prior to employment with ARI, he held research assistant positions with the U.S. Army Research Laboratory and the Virginia Modeling Analysis and Simulation Center (VMASC). Dr. Spain has received recognition for his work in educational and human factors psychology at the Interservice/Industry Training, Education, and Simulation Conference (I/ITSEC) and from the Human Factors and Ergonomics Society.

David Traum

David Traum is a principal scientist at the Institute for Creative Technologies (ICT), and research faculty in the Computer Science Department, both at the University of Southern California. At ICT, Traum leads the Natural Language Dialogue Group (http://projects.ict.usc.edu/nld/group/). Traum's research focuses on dialogue communication between human and artificial agents. He has engaged in theoretical, implementational and empirical approaches to the problem, studying human-human natural language and multi-modal dialogue, as well as building a number of dialogue systems to communicate with human users. He has pioneered several research thrusts in computational dialogue modeling, including computational models of grounding (how common ground is established through conversation), the information state approach to dialogue, multiparty dialogue, and non-cooperative dialogue.

Traum is author of over 200 technical articles, is a founding editor of the *Journal Dialogue and Discourse* has chaired and served on many conference program committees, and is on the board and a past president of SIGDIAL, the international special interest group in discourse and dialogue. He earned his Ph.D. in computer science at University of Rochester in 1994.

Matthew Ventura

Matthew Ventura is a Senior Research Scientist in the College of Education at Florida State University. His research centers around the design of educational games targeted towards STEM, cognitive skills, and health. His work has resulted in numerous publications in the areas of assessment and educational psychology.

Eliane Stampfer Wiese

Eliane Stampfer Wiese is a graduate student in Human-Computer Interaction at Carnegie Mellon, advised by Kenneth R. Koedinger. Eliane is an Institute of Education Sciences Research Training Fellow and a member of the Pittsburgh Science of Learning Center. Eliane's main research interests are educational and cognitive psychology in the context of STEM learning. Her research has focused on feedback to support sense-making in intelligent tutoring systems.

Beverly Park Woolf

Beverly Park Woolf, Ph.D., Ed.D., is a Research Professor in the School of Computer Science, UMass-Amherst who develops intelligent tutors that model affective and cognitive student characteristics and combine cognitive analysis of learning with artificial intelligence, network technology and multimedia. These systems represent the knowledge taught, recognize learners' skills and behavior, use sensors and machine learning to model student affect, and adjust problems to help individual students. Dr. Woolf has developed tutors for education and industry and in a variety of disciplines (e.g., chemistry, psychology, physics, geology, art history, mathematics and economics). Some tutors enable students to pass standard exams at a 10% higher rate and one system is used by more than 150,000 students per semester across hundreds of colleges. Dr. Woolf published the book *Building Intelligent Interactive Tutors* along with over 200 articles. She is lead author on the NSF report *Roadmap to Education Technology* in which forty experts and visionaries identified the next big computing ideas for education technology and developed a vision of how technology can incorporate deeper knowledge about human cognition. Dr. Woolf has delivered keynote addresses, panels and tutorials in more than 20 foreign countries and is a fellow of the American Association of Artificial Intelligence.

R. Michael Young

R. Michael Young is a professor of computer science at North Carolina State University, where he leads the Liquid Narrative Research Group. He's the founder and director of the NCSU Digital Games Research Center. His work focuses on the computational modeling of interactive narrative, especially in the context of computer games and virtual worlds. He teaches courses on game development and interactive narrative.

In 2000, Michael received a CAREER Award from the US National Science Foundation. He has received awards for both outstanding teaching and outstanding activities in economic development. In 2010, Michael was awarded a GlaxoSmithKline Faculty Fellowship for Public Policy and Public Engagement. Michael is a senior member of IEEE and of AAAI and an ACM Distinguished Scientist.

Michael was co-founder of several conferences that are leading venues for publication of scientific advances in computer games. Michael was editor-in-chief of the *Journal of Game Development* from 2007 to 2008 and an associate editor of the ACM journal *Transactions on Interactive Intelligence Systems* from 2012 through 2013. He serves as an associate editor of the IEEE journal *Transaction on Computational Intelligence and AI in Games* and the journal *Advances in Cognitive Systems*.

Instructional Management Advisory Board Members

The editors of *Design Recommendations for Intelligent Tutoring Systems: Volume 2 - Instructional Management* would like to recognize the contributions of instructional management advisory board members whose contributions to this effort were not in the authoring of book chapters, but provided motivations and focus to the development of this volume.

Scott Douglass
Dr. Scott A. Douglass is a Senior Research Psychologist with the 711/HPW Cognitive Models and Agents Branch (RHAC), US Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio. He holds a Ph.D. (2007) in cognitive psychology from Carnegie Mellon University. Working with John R. Anderson at CMU, he acquired expertise in cognitive architectures, eye tracking systems, and the modeling and simulation of complex situated cognitive processes. His research interests include large-scale cognitive modeling in event-driven computer-based systems, complex event processing, artificial intelligence, knowledge representation, multi-formalism modeling, and automated training aids. He is a member of the Society for Modeling and Simulation International (SCS). He can be reached at scott.douglass@wpafb.af.mil.

ACRONYM LIST

3D	three-dimensional
AAR	After Action Review
ABC	Airway-Breathing-Circulation
ABCDEF	airways, breathing, circulation, drugs, endocrine/electrolyte, fever, and general
ACPS	adaptive content and process scaffolding
ADDIE	Analysis, Design, Development, Implementation, and Evaluation
ADL	Advanced Distributed Learning
AI	artificial intelligenc
AIL	Add Incorrect Link
AL	Add Link
ALMA	A Layered Model of Affect
APS	adaptive process scaffolding
ARA	Acquiring Research Acumen
ARL	U.S. Army Research Laboratory
ASAT	AutoTutor Authoring Tools
ATWS	AutoTutor Web Service
BSCS	Brief Self Control Scale
CBLEs	computer-based learning environments
CDT	Component Display Theory
CITSRD	Center for Intelligent Tutoring Systems Research & Development
Con-G	Control group
CRC	current relevant contribution

CT	Cognitive Tutor®
CTAT	Cognitive Tutor Authoring Tool
CV	control-value
DDA	Dynamic Difficulty Adjustment
DOD	U.S. Department of Defense
DP	Deteriorating Patient
ECA	embodied conversational agent
EEG	electroencephalogram
EL	e-learning without support of self-regulation
EL+IMP	e-learning with self-metacognitive questioning
eMAP	engine for Managing Adaptive Pedagogy
EMG	electromyography
EMOTE	Embodied-Perceptive Tutors for Empathy-Based Learning
ERL	Externally Regulated Learning
ES	error searching
EMT	expectation and misconception tailored
FIT	Framework for Instructional Technology
Full-G	Full Support group
GAP	game-based assessment of persistence
GIFT	Generalized Intelligent Framework for Tutoring
GUI	graphical user interface
HRED	Human Research and Engineering Directorate
IES	U.S. Institute for Education Sciences
INSPIRE	intelligent, nurturant, Socratic, progressive, indirect, reflective, and encouraging
IIS	Institute for Intelligent Systems
IR	instructed reappraisal

ITS	intelligent tutoring system
KC	Knowledge Construction
KCs	knowledge components
LBT	learning by teaching
LCC	Learner Characteristic Curve
LITE	Learning in Intelligent Tutoring Environments
LMSs	Learning management systems
LPs	Learning Progressions
LSA	Latent Semantic Analysis
MAP	Measures of Academic Progress
MDPs	Markov decision processes
MKO	more knowledgeable other
NFC	need for cognition
NLP	natural language processing
NP	Newton's Playground
NPC	non-player character
NPV	Net Present Value
NS	no scaffolding
NSF	U.S. National Science Foundation
NWEA	Northwestern Education Association
OELEs	open-ended learning environments
OMB	U.S. Office of Management and Budget
OLMs	open learner models
Q	Quiz
R	Read
R&D	research and development

RIL	Remove Incorrect Link
RL	Remove Link
ROI	Return on Investment
RPS	Rock-Paper-Scissors
SASO	Stability and Support Operation
SE	Solution Evaluation
SE-G	Solution Evaluation group
SETs	Student Engagement Techniques
SLINQ	Science Learning by Inquiry
SME	subject matter expert
SPIKES	setting, perception, invitation, knowledge, empathy, summary and strategies
SQL	Structured Query Language
SRL	self-regulated learning
STEM	Science, Technology, Engineering and Mathematics
SZPD	specific ZPD
TAT	teachable agent trialogue
TERS	transfer effectiveness ratios
TLCTS	Tactical Language and Culture Training System
TRADEM	Tools for Rapid Automated Development of Expert Models
TRG	Tutoring Research Group
TVS	target variable strategy
VLT	vicarious learning trialogue
WTF	Without Thinking Fastidiously
XAI	explainable artificial intelligence
XML	extensible markup language
xPST	eXtensible Problem-Specific Tutor

ZPD   Zone of Proximal Development

INDEX

382

384

Design Recommendations for Intelligent Tutoring Systems

Volume 2
Instructional Management

Design Recommendations for Intelligent Tutoring Systems explores the impact of intelligent tutoring system design on education and training. Specifically, this volume examines "Instructional Management" techniques, strategies and tactics, and identifies best practices, emerging concepts and future needs to promote efficient and effective adaptive tutoring solutions. Design recommendations include current, projected, and emerging capabilities within the Generalized Intelligent Framework for Tutoring (GIFT), a modular, service-oriented architecture developed to promote simplified authoring, reuse, standardization, automated instructional management and analysis of tutoring technologies.

About the Editors:

- **Dr. Robert Sottilare** is the Chief Scientist at the Army Research Laboratory's SFC Paul Ray Smith Simulation & Training Technology Center. He leads adaptive tutoring research within ARL's Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT).

- **Dr. Arthur Graesser** is a professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis and is a Senior Research Fellow in the Department of Education at the University of Oxford.

- **Dr. Xiangen Hu** is a professor in the Department of Psychology at The University of Memphis and visiting professor at Central China Normal University. Dr. Hu received his Master (applied mathematics) from Huazhong University of Science & Technology, Master (Social Sciences) and Ph.D. (Cognitive Sciences) from the University of California, Irvine

- **Dr. Benjamin Goldberg** is an adaptive tutoring scientist at the U.S. Army Research Laboratory's SFC Paul Ray Smith Simulation & Training Technology Center. The focus of his research is adaptive tutoring, instructional strategies and their effect on learning. Dr. Goldberg is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT).

A Volume in the Adaptive Tutoring Series