

PART V

Multimedia Learning in Advanced Computer-Based Contexts

PROOF

PROOF

29 Multimedia Learning with Intelligent Tutoring Systems

Benjamin D. Nye
University of Memphis

Arthur C. Graesser
University of Memphis

Xiangen Hu
University of Memphis

Abstract

Intelligent tutoring systems (ITS) are a form of computer training that attempt to simulate both human tutors and ideal tutoring strategies. Learning gains from ITS environments have been promising, showing 0.4–1.1 standard deviations of improvement in learning over traditional instruction and the reading of texts. These systems show advantages in learning over conventional computer-based instruction, which in turn show improvements over the presentation of static material. This chapter reviews highlights of these ITS advances, particularly from the perspective of multimedia. Different types of multimedia have been integrated with these tutoring systems, including materials in multiple modalities, animation, interactive simulation, and conversational agents. The chapter identifies forms of multimedia implemented in tutoring systems driven by problem solving, simulations, natural language, games, and embodied interaction. Although these types of ITS routinely incorporate various multimedia, there is not a large body of research that assesses the incremental impact of particular multimedia features on learning or that segregates the contributions of the multimedia per se versus the intelligent mechanisms of ITS. Available evidence suggests that the conceptual content that is adaptively presented and interactively experienced by the student is more important than the type of media that presents the content. It is important to align media to the knowledge, skills, and strategies targeted by ITS and to resist media features that add irrelevant activities that distract from the serious educational content. More research is needed to determine how feedback, interactivity, attention focus features, hints, and other intelligent scaffolding are optimally coordinated with multimedia.

Introduction

Intelligent tutoring systems are a form of computer training that attempt to emulate aspects of personalized, one-on-one human tutoring as well as ideal tutoring strategies (Graesser, Conley, & Olney, 2012; Woolf, 2009). Learning gains from tutoring systems have been very promising, with recent meta-analyses and reviews reporting 0.4–1.1 standard deviations of improvement over traditional instruction (Dodds & Fletcher 2004; Graesser, Conley, & Olney, 2012; Steenbergen-Hu & Cooper, in press; VanLehn, 2011). Tutoring is a multifaceted process that requires communication, domain expertise, understanding of a student's knowledge, pedagogical strategies, and social interaction. The complexity of human tutoring has resulted in ITS applications specializing in different learning processes as well as content domains. This specialization drives tutoring systems to use multimedia in particular ways.

Tutoring systems employ a variety of modalities and media to accomplish their goals. Natural language tutoring systems, such as AutoTutor (Graesser et al., 2012), provide tutoring in a conversational manner by means of an agent (talking head) that engages in a dialogue with the student face to face. Figure 29.1 shows an example of an AutoTutor interface. AutoTutor has an agent that converses in natural language and that directs the interaction with other media, such as diagrams, questions, and conversation turns. The face is a universal interface that is familiar to the student, so there is no need to train the learner how to interpret the agents. Agents can assist the student in understanding and using other media, such as diagrams and embedded simulations in a virtual world (Graesser, Chipman, Haynes, & Olney, 2005; Johnson, Rickel, & Lester, 2000). Problem-based tutors for mathematics, such as ASSISTMents (Heffernan, Koedinger, & Razzaq, 2008), ALEKS (Hu et al., 2012), and the Cognitive Tutor (Ritter, Anderson, Koedinger, & Corbett, 2007), typically are confined to text and graphic images but use advanced artificial intelligence algorithms to generate problems, hints, and feedback. Simulation-based tutoring systems embed students in responsive dynamic environments rather than present content verbally and graphically alone (Fournier-Viger, Nkambou, & Mayers, 2008). Game-based tutoring systems in virtual worlds (Johnson, 2010) and some physically embodied robotic tutoring systems have very sophisticated approaches to using multimedia. Quite clearly, there is a diversity of interconnections between these automated tutoring systems and multimedia.

A systematic analysis of the links between ITS and multimedia is pursued in this chapter. First, we identify the multimedia elements that are used by different ITS applications. While most intelligent tutoring architectures present multimedia content, they do so somewhat differently. The following section describes tutoring designs that differ in their use of multimedia, with a representative example of each type. Second, we note that the ITS research

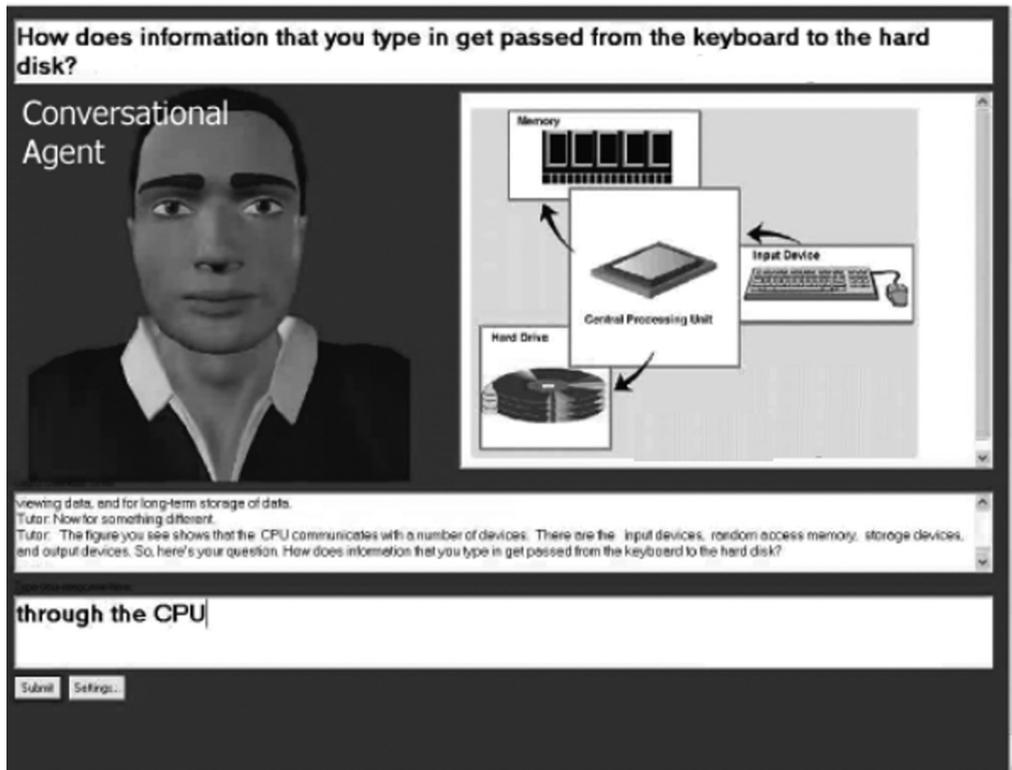


Figure 29.1. *AutoTutor natural language interface.*

community has placed significant emphasis on studying the impact of different multimedia designs on the effectiveness of learning and summarize findings from ITS research related to multimedia learning principles. Third, we discuss the implications of ITS findings for cognitive theory and instructional design. Finally, we identify some limitations of existing research and some directions for the future.

Examples of Multimedia Learning in ITS

Although ITS applications often have similar high-level behaviors, such as hints and scaffolding the difficulty of problems (VanLehn, 2006), they communicate with learners through a wide range of multimedia artifacts. An ITS is fundamentally interactive at a fine-grained level, so the range of media is virtually unlimited. Table 29.1 displays different presentation media used by ITS applications based on the sensory modalities through which they communicate and the presentation format that displays the information. For each combination, the table notes specific media formats that

Table 29.1. *Presentation media used by intelligent tutoring systems*

Presentation format	Sensory modality			
	Sight	Sound	Sight + sound	Touch
Verbal	Text	Voice	Text + voice	Braille
Nonverbal static	Graphic	Sound		Physical movement
Nonverbal dynamic	Animation	–	Simulation, video, or virtual world	Haptic feedback

have been used by tutoring systems. Tutoring systems use sight, sound, and touch as sensory modalities, with some media combining sight and sound (e.g., video). Presentation formats are broken down into three types: verbal, nonverbal (static), and nonverbal (dynamic). The presentation formats in Table 29.1 are generalizations of the distinctions among verbal, static graphics, and dynamic graphics framed by Mayer (2009). While multimedia learning research does not typically focus on nonverbal auditory information or touch modalities, some ITS environments do emphasize these formats and are worth discussing.

Verbal material is a common mode of communication in most tutoring systems. The tutors typically present verbal content using text, voice, or redundant text and voice, but a specialized Braille interface has also been developed (Kalra et al., 2009). Graphics and pictures are also quite common. Graphical highlighting of errors or sound effects to signal errors are less frequent, but not uncommon (Ritter et al., 2007; Suraweera & Mitrovic, 2004). Sound effects also deliver information in simulation-based learning where sounds carry contextual meaning, such as gunfire in a military simulation (Silverman et al., 2012). Systems that interact using nonverbal touch are uncommon and are typically designed for specialized purposes, such as disability accommodation (Kalra et al., 2009) or training in physical skills (Shih, 2012). Dynamic media are quite common in tutoring systems, which frequently include animated agents, videos, and virtual worlds (Aleven, McLaren, Sewall, & Koedinger, 2009; Biswas et al., 2010; Graesser, Jeon, & Dufty, 2008; Hu & Graesser, 2004; Johnson, 2010; Rowe, Shores, Mott, & Lester, 2011).

Table 29.2 presents five representative ITS designs and indicates typical user presentation media for each. This set of representative designs is not exhaustive, nor is it exclusive. While this table reflects common designs, tutoring systems that break this mold are not uncommon. For example, Cognitive Tutor Algebra fits the typical problem-solving tutor mold (Ritter et al., 2007). However, SimStudent, a pedagogical agent, supports tutoring by using the same underlying Cognitive Tutor models (Matsuda et al., in press). Table 29.2 is intended to be representative of typical designs rather than to imply that ITS applications are restricted to these archetypes.

Table 29.2. *Typical media of representative ITS designs*

	Type of tutor				
	Problem-solving	Simulation-based	Natural language	Game-based	Embodied / tangible
Text	X	X	X	X	X
Voice		X	X	X	X
Sound effects		X		X	
Graphics/pictures	X	X	X	X	
Animated agents			X	X	
Video/cut-scenes				X	
Virtual world		X		X	
Motion/haptic					X

The remainder of this section expands on Table 29.2 by presenting examples of these five tutor categories: problem-solving, natural language, simulation-based, game-based, and embodied. For each example, we briefly describe the common design elements, implementation characteristics, and empirical evaluations.

Problem-Solving ITS: Cognitive Tutor for Algebra

Problem-solving tutors organize tutoring activities around well-formed tasks, such as solving a math problem, writing a small code snippet, or selecting moves for end-game chess. Problem-solving tutors typically use a mixture of text and graphic media. The Cognitive Tutor is a well-established tutoring architecture based on the ACT-R theory of cognition that focuses on stepwise problem solving (Anderson, Corbett, Koedinger, & Pelletier, 1995; Ritter et al., 2007). The Cognitive Tutor family of ITS has more than 500,000 users and provides tutoring on topics such as algebra, geometry, and programming languages. Cognitive Tutor for Algebra I has been evaluated and designated as effective by the What Works Clearinghouse (U.S. Department of Education, Institute of Education Sciences, 2009). Since the development of Cognitive Tutor started approximately 30 years ago, Cognitive Tutor's design has had an influence on the design of later tutoring systems. Other major problem-solving tutors that focus primarily on step-based problem solving, such as ASSISTments (Heffernan et al., 2008) and Andes (VanLehn et al., 2005), present media in a similar fashion. Examples of the Cognitive Tutor interfaces can be found at the project Web site (www.carnegielearning.com, 2013).

Cognitive Tutor for Algebra is packaged as part of a full-year blended curriculum that includes an accompanying textbook. In many ways, the content displayed by the tutoring system resembles the combination of text and graphics that would be present in a math text. However, multiple principles

of representation and interaction are incorporated in the materials through text, graphics, highlighting, feedback, and other features of interactive multimedia. For example, students interact with the graphs and tables through actions (such as tagging spots and moving points), rather than the computer simply presenting the graphs and asking questions. Students sometimes fill out tables and justify steps in problem solving by selecting axioms or formulas from a menu of options.

A full suite of interactive multimedia facilities are incorporated into these problem-solving tutors, but there has not been a comprehensive and systematic investigation of which multimedia features have added value over and above the ACT-R components that are directly tied to problem solving. Nevertheless, some research has identified human-computer interface characteristics that have an impact on learning. A few examples are noteworthy. Anderson and Gluck (2001) conducted some studies in which they collected eye movement data while students solved problems on the algebra tutor. They discovered that the students rarely looked at the hint box that contained important suggestions or questions on what to do next. The interface needed to be modified to make the hints more prominent. As another example, the Cognitive Tutor sometimes provides an open-learner model called the “skill-o-meter,” which enables students to view the tutoring system’s estimates of their knowledge and competencies (Ritter et al., 2007). This gives learners some insight into their strengths and weaknesses. Students do pay attention to the skill-o-meter feedback, but empirical studies need to be conducted to explore the extent to which users take advantage of this feedback to improve their learning. As yet another example, problem-specific tutoring includes verbal hints (such as the next goal to be accomplished or a piece of useful information, e.g., “You know that the expression represents the savings”), verbal feedback (e.g., “That’s not quite right”), and graphical feedback (e.g., error highlighting). This feedback is mixed-initiative, in the sense that the user may request help or the system may automatically provide help. Significant debate exists over the appropriate balance of initiative for problem-solving tutors. Available data indicate that most students fail to activate ITS help functions even when stuck (Roll, Alevan, McLaren, & Koedinger, 2011), whereas some students “game the system” by overusing the help functions and relying on these help facilities to solve the problem without much deep thinking (Baker, Corbett, Roll, & Koedinger, 2008).

Simulation-Based ITS: CanadarmTutor

Simulation-based tutors also have a long track record, beginning with work as early as Lesgold, Lajoie, Bunzo, and Eggen’s (1988) presentation of the Sherlock tutoring system to help technicians learn to troubleshoot electronics. Simulation-based ITS remain popular, particularly for tasks that are tightly linked to an operational environment, such as power plants,

aviation, or military exercises. Tutors linked to an operational environment are sometimes referred to as *situated tutors* (Schatz, Oakes, Folsom-Kovarik, & Dolletski-Lazar, 2012). Media used by simulation-based tutors typically include text and graphics, with a larger emphasis on graphics than traditional problem-solving tutors. Sound effects may be employed if they provide important information about the domain environment. In virtual worlds, the student controls an avatar and uses realistic controls for machinery or scientific systems (Biswas et al., 2010; Graesser, McNamara, & VanLehn, 2005; Johnson et al., 2000).

As one example, the CanadarmTutor was designed to tutor operators of the Canadarm II robotic arm on the International Space Station using a simulation-based methodology (Fournier-Viger et al., 2008). Controlling Canadarm is difficult because the arm has 7 degrees of freedom and must move through a complex three-dimensional (3D) space where a small failure can have catastrophic consequences. Based on the paths for Canadarm selected by the student, the arm moves through a 3D virtual environment that is displayed through three monitor views. In addition to planning the movements for the arm, the learner must select and control 10 cameras that can be displayed on the monitors. These camera views present the actual environmental information available during the operation of the real-life Canadarm system. Text is available to provide feedback or suggest tasks for the learner, but the simulation dynamics provide the core learning material for the tutoring system. Examples of the CanadarmTutor interface are available in image and video form on the project Web site (www.philippe-fournier-viger.com/canadarmtutor, 2013).

Systems like CanadarmTutor are major technological achievements. However, it is difficult to identify controlled studies that compare the effectiveness of training by means of sophisticated simulation tutors with that of alternative training that attempts to control for content delivery or conventional training curricula. Studies are extremely sparse on *lesion* experiments that compare different versions of the ITS simulation that manipulate the presence or absence of particular multimedia features. There are many features to consider, so there would be many conditions to compare in a factorial design. Such tests are impractical for interdisciplinary design and delivery teams that have deadlines in their milestones.

Natural Language ITS: AutoTutor

Learning environments with conversational agents guide the interaction with the learner, suggest what the learner should do, adaptively respond to the learner's natural language, and interact with other agents to model ideal behavior, strategies, reflections, and social interactions (Graesser et al., 2008; Graesser et al., 2012; Millis et al., 2011). Some agents generate humanlike speech, gestures, body movements, and facial expressions, as exemplified by

Betty's Brain (Biswas et al., 2010), Tactical Language and Culture System (Johnson, 2010), iSTART (McNamara, O'Reilly, Best, & Ozuru, 2006), and My Science Tutor (Ward et al., 2011). Systems like AutoTutor and Why-Atlas can interpret the natural language of the human that is generated in spoken or typed channels and can respond adaptively to what the student expresses (Graesser et al., 2012; D'Mello, Dowell, & Graesser, 2011; VanLehn et al., 2007).

We focus on AutoTutor in part because there has been a systematic attempt to identify components of the complex system that are responsible for improving learning. Before turning to the learning gains, it is important that we convey some points about the mechanism. AutoTutor's dialogues are organized around difficult questions and problems whose answers require reasoning and giving explanations. For example, the following is a question on Newtonian physics:

If a lightweight car and a massive truck have a head-on collision, upon which vehicle is the impact force greater? Which vehicle undergoes the greater change in its motion, and why?

This question requires the learner to construct an ideal answer in approximately three to seven sentences and to exhibit reasoning in natural language. The dialogue involved in one of these challenging questions typically requires dozens of conversational turns between AutoTutor and the student. The structure of the dialogue in AutoTutor attempts to simulate that of human tutors and ideal tutors. AutoTutor's dialogue is centered around a set of expectations and misconceptions. AutoTutor keeps the dialogue on track because it is always comparing what the student says with these anticipated inputs (i.e., the expectations and misconceptions in the curriculum script). Pattern-matching operations and pattern completion mechanisms drive the comparison. These matching and completion operations are based on latent semantic analysis (Landauer, McNamara, Dennis, & Kintsch, 2007) and symbolic interpretation algorithms (Graesser et al., 2012). AutoTutor cannot interpret student contributions that have no matches to content in the curriculum script. This, of course, limits true mixed-initiative dialogue.

The learning gains produced by AutoTutor have been evaluated in more than 20 experiments conducted during the past 16 years. Assessments of learning gains have shown effect sizes of approximately 0.8 standard deviation in the areas of computer literacy (Graesser et al., 2004) and Newtonian physics (VanLehn et al., 2007). AutoTutor's learning gains have varied between 0 and 2.1 sigma (a mean of 0.8), depending on the learning performance measure, the comparison condition, the subject matter, and the version of AutoTutor. Approximately a dozen measures of learning have been collected in these assessments on the topics of computer literacy and physics, including (1) multiple choice questions on shallow knowledge that tap definitions, facts, and properties of concepts; (2) multiple choice questions on deep

knowledge that tap causal reasoning, justifications of claims, and functional underpinnings of procedures; (3) essay quality when students attempt to answer challenging problems; (4) a cloze task in which subjects fill in missing words of texts that articulate explanatory reasoning on the subject matter; and (5) performance on problem-solving tasks.

There have been some attempts to examine how the different components of AutoTutor differentially influence learning (Graesser et al., 2004; Graesser et al., 2008; VanLehn et al., 2007). The assessment methodology tested the differences between the normal conversational AutoTutor's effect size on learning gains against the learning produced by different comparison conditions. The conversational AutoTutor had (1) a 0.80 effect size (σ) compared with pre-tests, a control that entailed reading a textbook, or doing nothing; (2) a 0.22 σ compared with reading textbook segments directly relevant to the AutoTutor problems; (3) a 0.07 σ compared with reading a script that succinctly answered the questions posed by AutoTutor; (4) a 0.13 σ compared with AutoTutor presenting speech acts in print instead of the talking head; and (5) a 0.08 σ compared with expert human tutors in computer-mediated conversation. Compared with the conversational version, there was an increment of 0.20 σ that added interactive 3D simulations to the AutoTutor conversations. That is, the best system had a conversation with the agent that added the sophisticated multimedia of interactive simulations. From the standpoint of the modality of how students express themselves in natural language, there has been no difference in learning gains between speech and typing (D'Mello et al., 2011).

The aforementioned comparison and lesion studies suggest that more economical systems without agents might be almost as effective in promoting learning even though the results show a modest advantage of systems with animated agents. Presenting the right content at the right time has a more robust impact on learning than the talking-head and conversational features. However, it is important to consider other factors in assessments of learning environments, such as motivation and individual differences. Systems with agents may keep some students more motivated and optimize persistence. Some students may better learn from oral conversation than reading printed text. Future research needs to systematically investigate the role of motivation and individual differences in these comparison studies with agents.

Game-Based ITS: Crystal Island

Game-based tutoring systems make use of the widest range of media to present content, especially those that combine elements from natural language tutors and simulation-based tutors. While game-based tutoring systems are commonly associated with 2D or 3D virtual worlds (Mavrikis, Gutierrez-Santos, Geraniou, & Noss, 2012; Rowe et al., 2011), they also include game show formats or simple casual games (Lomas et al., 2012; Matsuda

et al., in press). Nearly all games include text, graphics, and sound effects, whereas games based on virtual worlds often include animated agents and videos. Sometimes simple ITS games, such as the SimStudent quiz show game (Matsuda et al., in press), include elements of problem-solving tutors. Games based on 3D worlds often deliver content through scripted or branching dialogue, whereas natural language tutoring with sophisticated computational linguistics is seldom provided. One notable exception is Operation ARA (Acquiring Research Acumen), a 3D game based on AutoTutor that employs natural language “trialogs” to support student’s critical thinking skills on scientific reasoning (Halpern et al., 2012; Millis et al., 2011). A trialog is a conversation between a human and two agents (a tutor agent and a student peer agent). The trialogs have been shown to significantly improve scientific reasoning beyond reading comparison texts (Kopp, Britt, Millis, & Graesser, 2012). However, the impact of the coherent narrative has not yet been assessed.

Crystal Island is a narrative-based intelligent learning environment for teaching scientific inquiry and K–12 science topics such as genetics and microbiology (Rowe et al., 2011). Users navigate Crystal Island through an avatar to select dialogue choices with other agents and interact with objects, such as computer terminals. A video of this interface is presented on the project Web site (www.intellimedia.ncsu.edu/ci8.html, 2013). While many animated agents exist in the virtual world, their dialogue is directed by an interconnected narrative structure that is affected by both the dialogue and the results of in-game experiments conducted by the student (Rowe et al., 2011). Rather than a single tutoring agent, the virtual world itself provides adaptive help using different animated agents. Crystal Islands has many multimedia and intelligent components, but the incremental value-added impact of particular components on learning has rarely been assessed. There is no substantial evidence that Crystal Island is more effective than static instructional materials. Available evidence suggests that the coherent narrative may detract from learning because it competes with the processing of serious educational content (Adams, Mayer, MacNamara, Koenig, & Wainess, 2012). The narrative component may play a greater role in optimizing motivation and persistence as opposed to efficient learning.

Tangible and Embodied ITS: Braille Tutor

Embodied tutors systems have been designed for disparate reasons, including disability accommodation (Kalra et al., 2009), providing a tangible animated agent (Bickmore, Schulman, & Vardoulakis, in press; Wang, Young, & Jang, 2009), and tutoring of physical activities (Shih, 2012). The unifying feature of these systems is that they communicate with the user by means of physical movement (e.g., a robot moving), respond to user touch, or provide haptic feedback (e.g., vibration). These systems also commonly use verbal

communication (text or voice). Some use sound or graphics, but this depends greatly on the specific tutor.

The Braille Reading Tutor is an example of a tactile tutor for supporting learners with disabilities and uses a custom-designed e-Slate interface (Kalra et al., 2009). It was designed as a low-cost ITS for learning Braille in a low-resource context, such as rural India. Tutoring follows a scaffolding methodology, initially providing a high level of support and fading the scaffold as the learner progresses. Learners use a stylus to write Braille on a special e-Slate, which provides voice feedback when they have made a mistake on a character. Additionally, a set of six press buttons represent a single large Braille cell where users can physically manipulate a character. Pictures of this e-Slate are available on the project Web site (www.techbridgeworld.org/brailletutor, 2013). For young learners, “friendly” audio feedback and a recording of a local teacher’s voice provide corrective feedback while maintaining student motivation (Kalra et al., 2009). While pilot studies have been performed, this system has not been adequately tested on a reasonable sample of students to evaluate its effectiveness.

Findings on Multimedia Learning with ITS

Most ITS researchers conduct empirical studies that assess the impact on learning as the systems are built. This is an important part of the development process because these tutoring systems are expensive to build. This section first summarizes benchmark studies on major tutoring systems in order to convey the overall effectiveness of ITS compared with traditional classroom instruction and learning materials such as textbooks. We subsequently report results from deconstructed and parameterized study designs that provide insight into which features of tutoring designs contribute to learning gains.

Learning Gains and Benchmark Studies for Tutoring Systems

There is considerable evidence that, as a group, ITS significantly improve learning. VanLehn’s (2011) estimate from a meta-analysis was an effect size of $d = 0.76$ for ITS when compared with static materials that entailed no tutoring. In contrast, human tutoring (expert and nonexpert grouped together) provided an effect size of $d = 0.79$, essentially the same. Graesser, Conley, and Olney (2012) reported estimates between 0.8 and 1.0 sigma over various controls, whereas Dodds and Fletcher (2004) gave an estimate as high as 1.08 and Steenbergen-Hu and Cooper (in press) one as low as 0.37.

Problem-solving tutors have demonstrated strong effectiveness in benchmark studies. The Cognitive Tutor Algebra I blended curriculum produced performance gains of approximately 0.3 sigma over traditional math curricula on standardized tests (Ritter et al., 2007). Tests specifically targeting

the problem-solving and multiple representations (a focus of Cognitive Tutor) showed effect sizes between 0.7 and 1.2 sigma (Ritter et al., 2007). ASSISTments, a Web-based homework ITS for math, demonstrated learning gains of 0.23 sigma on MCAS (Massachusetts Comprehensive Assessment System) items in a classroom field study that compared the intervention against traditional instruction in a school that requested ASSISTments but was not yet using it (Koedinger, McLaughlin, & Heffernan, 2010). ASSISTments showed gains of approximately 1.0 sigma from the Web-based homework with tutoring in an experimental study (Singh et al., 2011). Students using Andes, a physics problem-solving tutor, demonstrated strong learning gains across multiple studies (average of $d = 0.61$) over students who read static materials with no tutoring under experimental conditions (VanLehn et al., 2005). Finally, students trained in math fluency by Wayang Outpost showed significant learning gains when compared with control groups receiving traditional instruction only; although a standardized effect size was not reported, the post-test mean for the Wayang condition was approximately 18% higher than that of the control (Arroyo, Royer, & Woolf, 2011).

Natural language tutoring has similarly yielded noticeable improvements in learning. AutoTutor showed learning gains of 0.8 sigma compared with reading a textbook for an equivalent amount of time (Graesser, Conley, & Olney, 2012), as mentioned earlier in this chapter. Evaluations showed that Operation ARA, a dialog-based game based on AutoTutor, produced learning gains ($d = 1.4$) on a test of scientific research skills when compared against learners who received no instruction (Halpern et al., 2012). GuruTutor produced strong learning gains ($d = 0.72$) for biology concepts versus controls in a classroom and found that these gains persisted on a delayed post-test (Olney et al., 2012).

Simulation-based tutors are benchmarked on the basis of the amount of time or instructor support needed to complete a given level of training, possibly due to lack of standardized assessment measures for their material. Schatz et al. (2012) reviewed 86 situated tutoring systems and found that their evaluations often considered the efficiency to deliver materials rather than learning gains. On these measures, simulation-based tutors have been highly effective. Evaluation of Sherlock, an avionics repair tutor, showed that 20–25 hours of training with the tutor provided skills similar to those obtained after four years of on-the-job training (Lesgold et al., 1988). TAO-ITS, a naval surface warfare tutor, reduced the number of instructors from a ratio of 2 students per instructor to a single instructor for each class of approximately 40 students (Stottler & Panichas, 2006). In general, simulation-based tutors appear to be significantly cheaper and faster than traditional training methods, but the evidence for equivalent or enhanced learning is not firmly established in most cases.

Game-based ITS have received substantial attention in recent years because of the need to enhance motivation in systems that target deep learning rather

than shallow learning. Deep learning is challenging to students, rife with confusion and possible frustration, so motivation is an important component to keep them engaged (D'Mello, Lehman, Pekrun, & Graesser, in press). The number of systematic studies that assess learning is limited, but results are evolving (O'Neil & Perez, 2008). Studies of Crystal Island (Rowe et al., 2011), the Tactical Language and Cultural Training System (Johnson, 2010), and BiLat (Hays et al., 2009) have shown significant learning gains using pre-test/post-test designs, but have not reported effect sizes compared with benchmarks (e.g., human tutors, static media). Operation ARA provided strong learning gains, as noted previously (Halpern et al., 2012; Millis et al., 2011). In contrast, students learning with the iSTART reading tutor gave better self-explanations and used more content words in a non-game condition than students in a game condition (Jackson, Dempsey, & McNamara, 2011); however, students in a multi-week study produced longer self-explanations in the game condition without a drop in quality. This indicates a possible trade-off between learning efficiency and accumulated learning as a result of continued usage. Overall, simulation-based and game-based ITS could benefit from an increased focus on controlled benchmark assessments, which is part of a larger need for rigorous assessment of simulation and game-based learning technologies (National Research Council, 2010).

In summary, these benchmark studies show a few systems with strong evidence of effectiveness when compared with traditional approaches (e.g., static materials, human tutors, or traditional classroom instruction alone). Cognitive Tutor, Andes, AutoTutor, and ASSISTMents have demonstrated learning gains across multiple controlled studies (Graesser, Conley, & Olney, 2012; Koedinger, McLaughlin, & Heffernan, 2010; Ritter et al., 2007; VanLehn et al., 2005). Wayang Outpost and GuruTutor have demonstrated learning gains compared with benchmark classroom controls (Arroyo, Royer, & Woolf, 2011; Olney et al., 2012). Operation ARA, Crystal Island, the Tactical Language and Cultural Training System, and BiLat have demonstrated learning gains on pre-test/post-test designs but have yet to be compared with traditional methods or conditions that control for content (Halpern et al., 2012; Hays et al., 2009; Johnson, 2010; Rowe et al., 2011). These pre-test/post-test comparison designs were popular in research on early ITS prototypes, when resources for larger studies were not available. However, it is important to move forward and compare such systems with conditions that involve reading a textbook or static HTML pages for an equivalent amount of time. Sherlock, TAO-ITS, and other simulation-based systems have demonstrated higher learning efficiency (e.g., faster, fewer instructors, etc.), but no studies have assessed whether learning was different from or equivalent to that under comparison conditions.

The studies in this section testify to the value of ITS in promoting learning, but the role of multimedia in explaining the learning gains has not been determined. Is it multimedia, intelligent adaptive algorithms, substantive

content, feedback, interactivity, natural language, agents, or some other component of these systems that explains the learning gains? This is destined to be a pressing question for research in the future.

Basic Multimedia Principles in ITS

ITS researchers have occasionally explored some of the basic multimedia factors, such as modality, redundancy, extraneous processing, and social cues (Clark & Mayer, 2003; Mayer, 2009). The modality and redundancy principles have been tested by a number of ITS studies conducted with AutoTutor. For example, Graesser et al. (2003, 2004) reported no significant learning differences when they compared text interaction, voice without an animated agent, animated agent with text, and a talking animated agent. A later pair of studies involving AutoTutor examined the impact of learners expressing their verbal input by speech versus text but found no significant effects for input modality (D'Mello et al., 2011). These studies are consistent with the conclusion that it is the content that matters rather than the presentation or input modalities. Specific modalities have affordances that presumably are natural for particular types of substantive content, so it is important to delineate content versus modalities in future research. Further attention to the representation of content is also important. Schnotz and Bannert (2003) suggest that text offers more traction for conveying abstract relationships, while graphics excel at presenting specific instances and diagrams used to draw inferences. In this view, language and graphics contribute complementary information to a shared mental model for a task. Particularly for natural language tutors, the strengths of speech and graphics to iteratively enrich a learner's mental model is a potentially powerful area that requires further exploration.

Extraneous processing is a significant topic for ITS research to the extent that there are components that potentially add a cognitive load or detract from serious learning. Simulations, games, inquiry facilities, and other facilities may enhance motivation and presumably promote active construction of learning, but some may steer the student away from serious content that might be better mastered through direct instruction (Clark & Mayer, 2003). Tutoring systems have the capability to monitor student interaction intelligently, but the complexity of learning may be a serious challenge to students and be prone to inducing inattention, gaming the system, and other disengaged behaviors. Studies based on eye tracking (Anderson and Gluck, 2001; D'Mello, Olney, Williams, & Hays, 2012) as well as other channels of perception and interactions (Baker et al., 2010) have found such disengagement from deep learning. A study of ASSISTments and gaming behavior (in which students avoid learning by asking for help without thinking deeply) found that students tend to game the system on problems in which they have relatively low knowledge and end up learning almost nothing (Gong et al.,

2010). This study also revealed that the best predictor for gaming was the students themselves, indicating strong individual differences in the level of gaming.

Some ITS researchers have investigated social cues, particularly personalization of communication. Experiments with the CTAT Stoichiometry Tutor examined differences between polite and direct tutoring (McLaren, DeLeeuw, & Mayer, 2011). Politeness increased learning ($d = 0.64$) for low-knowledge students on an immediately following post-test, which persisted on a delayed test ($d = 0.50$). However, polite tutoring did not help high-knowledge learners, for whom there was a nonsignificant negative effect on learning gains compared students who received direct tutoring ($d = -0.58$). This result is consistent with AutoTutor versions that manipulated conversational styles that were emotionally neutral, polite, or confrontational (D'Mello & Graesser, 2012).

Advanced Multimedia Principles in ITS

Research on advanced learning environments (not just ITS) has investigated multimedia factors that have obvious links to deeper levels of cognition, such as feedback, learner control, generation of information, interactivity, reflection, animation, simulation, and affect (Mayer, 2009; Moreno & Mayer, 2005, 2007). ITS developers have often adopted many of these multimedia principles that are supported by cognitive research, but there is not an abundance of lesion studies that evaluate the incremental impact of particular components on learning. There have been lesion studies on the role of feedback in deep learning. Tutoring systems by definition provide feedback, but questions remain about when, what, and how much feedback to offer learners (Shute, 2008). The amount of information provided in the feedback may range from minimal (such as correct vs. incorrect) to an elaborate explanation of a student's misconceptions and associated corrections. Experiments with MetaTutor, an offshoot of AutoTutor, showed that metacognitive prompts (e.g., suggestions that learners self-assess their understanding) combined with immediate feedback on students' metacognitive judgments resulted in higher learning efficiency than prompts alone and produced larger overall gains ($d = 0.84$) when the system did not provide feedback (Azevedo et al., 2012). Research on ASSISTMents showed advantages for interactive tutoring ($d = 0.54$) over immediate error correction (Singh et al., 2011). These and other studies indicate significant gains for feedback in tutoring systems and also significant gains of ITS applications over basic error correction feedback (VanLehn, 2011).

One of the central questions in ITS assessment is whether learner control and interactivity with the tutor show incremental gains over content delivery or computer-based instruction with basic error correction feedback (VanLehn, 2011; VanLehn et al., 2007). The evidence leans toward added

value of learner control and interactivity, but the increment may substantially diminish to the extent that the system delivers relevant content at the right time or presents worked examples that display important reasoning. Of course, a major mission of tutors is to provide intelligent, adaptive information delivery in response to active learners, so it could be argued that such comparisons are part of what tutors do. One challenge with learner control is that students rarely ask questions, ask for help, or take the initiative in human tutoring (Graesser & Person, 1994) and computer learning environments (Alevan, Stahl, Schworm, Fischer, & Wallace, 2003; Graesser, McNamara, & VanLehn, 2005), unless they are gaming the system to avoid deep learning. Razzaq and Heffernan (2010) reported stronger learning gains when learners received hints only by request, but this finding was relevant only for students who tended to ask for many hints. VanLehn (2011) concluded that higher interactivity (e.g., more asking by the computer, less telling) has added value but hits a 1 sigma plateau with respect to learning gains, suggesting that further learning gains might be possible only with better targeting of the content rather than more interaction. A study comparing dialogue interactivity in AutoTutor showed that a mixture of intense dialogue on some problems and no dialogue in others yielded greater learning gains than intense dialogue only ($d = 0.46$), with students in these two conditions significantly outperforming those in a no-dialogue condition (Kopp et al., 2012).

Research on animation has generally had mixed success in promoting learning gains (Ainsworth, 2008; Hegarty, Kriz, & Cate, 2003). Animations run the risk of moving at a faster pace than the mind can keep up with. Interactive animations would be expected to be an improvement, and sometimes mental animations are the most powerful. Studies on AutoTutor on Newtonian physics showed a 0.2 increment from interactive animations in a 3D environment compared with a strictly conversational AutoTutor (Graesser et al., 2008). Students were able to manipulate parameters, such as the mass and speed of vehicles, and then observe how they move and sometimes collide with other entities. Unfortunately, most students do not interact in a strategically sensible manner that runs several simulations, manipulates parameters, and records the outcomes of the simulations. Instead, they run only one simulation, so their learning gains are near zero. Those who take the initiative to actively use the 3D systems with multiple runs do show learning, and these tend to be more knowledgeable students. Research on 3D animations has rarely been integrated with ITS, so that should be a direction for future research.

Implications for Cognitive Theory and Instructional Design

One central implication of the ITS research reported in this chapter is that the interactive content, rather than the multimedia format,

explains most of the learning gains. Multiple studies using different ITS have demonstrated a fairly consistent pattern of learning gains: interactive tutoring > computer-based tutoring with immediate feedback > static materials (Graesser, Conley, & Olney, 2012; Singh et al., 2011; VanLehn, 2011). A range of tutoring system designs (problem-solving, natural language, game-based) have evidence of learning gains between 0.4 1.1 standard deviations over static materials such as reading textbooks (Graesser, Conley, & Olney, 2012; VanLehn, 2011). ITS have shown gains across different types of assessments, improving performance on shallow knowledge, deep knowledge, and high-stakes standardized tests (Koedinger et al., 2010; Olney et al., 2012; Ritter et al., 2007). However, the greatest gains are for deeper learning and experimenter-designed tests over shallow learning and standardized tests. In summary, ITS have empirical support as a learning technology.

Not surprisingly, studies have confirmed that greater use and attention to knowledge, skills, and strategies of the subject matter correlate with higher learning gains (D'Mello et al., 2012; Koedinger et al., 2010; Roll et al., 2011). Off-task and gaming behaviors reduce the learning efficiency of tutoring systems (Baker et al., 2010, 2011). Attention to nongermane features, including some types of multimedia, runs the risk of diverting attention from important material. The key challenge is to make sure that multimedia components are aligned with the serious content and that they unveil information that the learner would not normally construct without the multimedia.

One successful principle of tutoring systems is to use step-based tutoring, where a tutor can provide feedback at any point during the student's problem-solving process rather than rigid error correction feedback on answers or solutions (VanLehn, 2011). The multimedia should facilitate important components of the timely feedback. The feedback ought to be adaptive to the student's knowledge and skill level rather than rigidly mechanical (like some agent voices in automobiles). The multimedia needs to direct the student's attention to the correct information among the complex array of material. It should signal what are correct or incorrect actions/decisions. It needs to support ITS delivery of information that provides deep explanations, corrections of misconceptions, and suggestions for self-regulated reasoning. It ought to elicit information from the student with hints rather than the mere delivery of information. All of these components need to be coordinated over time in a way that is motivating to the student. The dance between multimedia and ITS intelligent mechanisms is hardly simple. It requires a well-engineered mapping between multimedia components and ITS components.

Limitations and Future Directions

There are limitations of ITS research from the standpoint of multimedia design. These limitations should stimulate future research projects.

The first limitation is that ITS applications normally include a large suite of multimedia features. In some tutoring systems, the various features are motivated by cognitive and motivational principles of good multimedia design (Mayer, 2009). The good news is that the multimedia features are grounded in science. The disappointing news is that it is difficult to determine what multimedia components are partially responsible for the learning gains or student impressions of the system. Lesion studies need to be conducted that turn on or off each of the various components and track the consequences on learning and impressions. That would require many conditions. If there were 5 features (agents, highlighting, modality, etc.), that would require $2^5 = 32$ conditions. If 30 participants were run in each condition, that would require 960 participants. That is, of course, an onerous load on testing. However, if these ITS are available on the Web for free, such a participant sample is reasonable (Pardos, Dailey, & Heffernan, 2011).

A second limitation is related to the creation of a systematic alignment between multimedia features to the ITS components. This requires a detailed mapping that few if any tutoring systems are currently achieving. Suppose an ITS is adaptive to a particular student's profile, determines that a particular idea needs to be expressed by the student, and presents a hint to elicit the idea from the student. How should the multimedia interface deliver the hint? Should it be a printed verbal question in a window? Or a spoken question? Or a highlighted help icon on the interface? The design for this context is very different from that for a context in which a lost student is presented multimedia features to show the student a worked-out solution to a problem. Suppose the student is totally lost. Is it best to present a Khan Academy video or to have the student read a specific text that is tailored to his or her problem? An answer to this would require a profile of the student's cognitive inclinations. The science of intelligent computer-human interfaces would be needed to address this challenge.

A perfect alignment between ITS components and multimedia would, of course, be a major achievement. If that dream were achieved, the next challenge would be to determine whether it is the content or the multimedia that is responsible for learning gains or impressions. That is perhaps the most profound challenge and may even be unanswerable.

A third limitation lies in understanding how multimedia influences motivation or persistence in the ITS. The ideal litmus test is how much students use an ITS on their own, without curriculum requirements, money, or other external rewards. Multimedia would presumably increase engagement and intrinsic motivation for students to learn the difficult content that is the hallmark of ITS applications. There is a trade-off between learning and liking when it comes to difficult academic content. Perhaps multimedia or games could reduce or reverse the negative relationship between deep learning and the enjoyment of the learning experience.

Glossary

Benchmark study: A study that compares a learning technology against a control group using a traditional approach, such as static learning materials, classroom instruction, or human tutoring.

Game-based tutoring system: An intelligent tutoring system built into a game environment, where users learn while attempting to accomplish game objectives such as progressing through a 3D environment or increasing their score on a trivia game.

Intelligent tutoring system: A computer-based training system that provides personalized interaction with a user at a fine-grained level, attempting to emulate human tutoring and idealized tutoring strategies.

Natural language tutoring system: An intelligent tutoring system that receives language input (voice or text) and replies with natural language feedback, typically emulating the interaction between a human tutor and learner.

Problem-solving tutoring system: An intelligent tutoring system that organizes tutoring activities around well-formed tasks, such as solving a math problem. Typically provides step-based feedback and adaptive sequencing of problems based on performance.

Simulation-based tutoring system: An intelligent tutoring system that embeds students in a responsive environment that represents knowledge through the relationships and dynamics of the environment.

Situated tutoring system: A simulation-based tutoring system that embeds students into an operational environment, such as a flight simulator or power control system.

Tactile/embodied tutoring system: An intelligent tutoring system that uses a physical agent, such as a robotic agent, or requires tactile interaction, such as a haptic interface.

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*(1), 235–239.
- Ainsworth, S. (2008). How do animations influence learning? In D. Robinson & G. Schraw (Eds.), *Current perspectives on cognition, learning, and instruction: Recent innovations in educational technology that facilitate student learning* (pp 37–67). Charlotte, NC: Information Age.
- 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.

- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73*(3), 277–320.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*(2), 167–207.
- Anderson, J. R., & Gluck, K. (2001). What role do cognitive architectures play in intelligent tutoring systems? In D. Klahr & S. M. Carver (Eds.), *Cognition & instruction: Twenty-five years of progress* (pp. 227–262). Hillsdale, NJ: Lawrence Erlbaum.
- Arroyo, I., Royer, J. M., & Woolf, B. P. (2011). Using an intelligent tutor and math fluency training to improve math performance. *International Journal of Artificial Intelligence in Education, 21*(1), 135–152.
- Azevedo, R., et al. (2012). The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with MetaTutor. In S. A. Cerri, & B. Clancey (Eds.), *Proceedings of Intelligent Tutoring Systems (ITS) 2012* (pp. 256–261). Berlin: Springer.
- Baker, R. S., Corbett, A. T., Roll, I., & Koedinger, K.R. (2008) Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction, 18*(3), 287–314.
- 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*(2), 223–241.
- Baker, R. S., Goldstein, A. B., & Heffernan, N. T. (2011). Detecting learning moment-by-moment. *International Journal of Artificial Intelligence in Education, 21*(1), 5–25.
- Bickmore, T., Schulman, D., & Vardoulakis, L. (in press). Tinker: A relational agent museum guide. *Autonomous Agents and Multi-Agent Systems*.
- Biswas, G., Jeong, H., Kinnebrew, J., Sulcer, B., & Roscoe, R. (2010). Measuring self-regulated learning skills through social interactions in a teachable agent environment. *Research and Practice in Technology-Enhanced Learning, 5*(02), 123–152.
- Clark, R. C., & Mayer, R. E. (2003). *e-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. San Francisco: Jossey-Bass.
- 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), 1–17.
- D'Mello, S. K., & Graesser, A. C. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145–157.
- D'Mello, S. K., Lehman, B., Pekrun, R., & Graesser, A.C. (in press). Confusion can be beneficial for learning. *Learning and Instruction*.
- D'Mello, S. K., 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.
- Dodds, P., & Fletcher, J. D. (2004). Opportunities for new “smart” learning environments enabled by next-generation web capabilities. *Journal of Educational Multimedia and Hypermedia, 13*(4), 391–404.

- Fournier-Viger, P., Nkambou, R., & Mayers, A. (2008). Evaluating spatial representations and skills in a simulator-based tutoring system. *IEEE Transactions on Learning Technologies*, 1(1), 63–74.
- Gong, Y., Beck, J. E., Heffernan, N. T., & Forbes-Summers, E. (2010). The fine-grained impact of gaming (?) on learning. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of Intelligent Tutoring Systems (ITS) 2010* (pp. 194–203). Berlin: Springer.
- Graesser, A. C., Chipman, P., Haynes, B., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612–618.
- Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent tutoring systems. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), *APA educational psychology handbook*, vol. 3: *Application to learning and teaching* (pp. 451–473). Washington, DC: American Psychology Association.
- 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., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes*, 45(4–5), 298–322.
- 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, and Computers*, 36(2), 180–192.
- Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist*, 40(4), 225–234.
- Graesser, A. C., Moreno, K., Marineau, J., Adcock, A., Olney, A., & Person, N. (2003). AutoTutor improves deep learning of computer literacy: Is it the dialog or the talking head? In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Proceedings of Artificial Intelligence in Education (AIED) 2003* (pp. 47–54). Amsterdam: IOS Press.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, 31(1), 104–137.
- 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(2), 93–100.
- Hays, M., Lane, H. C., Auerbach, D., Core, M. G., Gomboc, D., & Rosenberg, M. (2009). Feedback specificity and the learning of intercultural communication skills. In V. Dimitrova, R. Mizoguchi, B. DuBoulay, & A. Graesser (Eds.), *Proceedings of Artificial Intelligence in Education (AIED) 2009* (pp. 391–398). Amsterdam: IOS Press.
- Heffernan, N. T., Koedinger, K. R., & Razzaq, L. (2008) Expanding the model-tracing architecture: A 3rd generation intelligent tutor for algebra symbolization. *International Journal of Artificial Intelligence in Education*, 18(2), 153–178.
- Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and Instruction*, 21(4), 209–249.
- Hu, X., Craig, S. D., Bargagliotti A. E., Graesser, A. C., Okwumabua, T., Anderson, C., Cheney, K. R., & Sterbinsky, A. (2012). The effects of a traditional and

- technology-based after-school program on 6th grade students' mathematics skills. *Journal of Computers in Mathematics and Science Teaching*, 31(1), 17–38.
- Hu, X., & Graesser, A. C. (2004). Human Use Regulatory Affairs Advisor (HURAA): Learning about research ethics with intelligent learning modules. *Behavior Research Methods, Instruments, and Computers*, 36(2), 241–249.
- Jackson, G. T., Dempsey, K. B., & McNamara, D. S. (2011). Short and long term benefits of enjoyment and learning within a serious game. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of Artificial Intelligence in Education (AIED) 2011* (pp. 139–146). Berlin: Springer.
- Johnson, W. L. (2010). Serious use of a serious game for language learning. *International Journal of Artificial Intelligence in Education*, 20(2), 175–195.
- Johnson, W. L., Rickel, J. W., and Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11(1), 47–78.
- Kalra, N., Lauwers, T., Dewey, D., Stepleton, T., & Dias, M. B. (2009). Design of a Braille writing tutor to combat illiteracy. *Information Systems Frontiers*, 11(2), 117–128.
- Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasiexperimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research*, 43(4), 489–510.
- Kopp, K. J., Britt, M. A., Millis, K., & Graesser, A. C. (2012). Improving the efficiency of dialogue in tutoring. *Learning and Instruction*, 22(5), 320–330.
- Landauer, T. K., McNamara, D. S., Dennis, S. E., & Kintsch, W. E. (2007). *Handbook of latent semantic analysis*. Hillsdale, NJ: Lawrence Erlbaum.
- Lesgold, A., Lajoie, S., Bunzo, M., & Eggan, G. (1988). Sherlock: A coached practice environment for an electronics troubleshooting job. In J. Larkin, R. Chabay, & C. Scheftic (Eds.), *Computer-assisted instruction and intelligent tutoring systems: Establishing communication and collaboration*. Hillsdale, NJ: Lawrence Erlbaum.
- Lomas, D., Stamper, J., Muller, R., Patel, K., & Koedinger, K. R. (2012). The effects of adaptive sequencing algorithms on player engagement within an online game. In S. A. Cerri & B. Clancey (Eds.), *Proceedings of Intelligent Tutoring Systems (ITS) 2012* (pp. 588–590). Berlin: Springer.
- Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., William, W. C., Stylianides, G. J., & Koedinger, K. R. (in press). Cognitive anatomy of tutor learning: Lessons learned with SimStudent. *Journal of Educational Psychology*.
- Mavrikis, M., Gutierrez-Santos, S., Geraniou, E., & Noss (2012). Design requirements, student perception indicators and validation metrics for intelligent exploratory learning environments. *Personal and Ubiquitous Computing*. Springer OnlineFirst. doi: 10.1007/s00779-012-0524-3, 1-16.
- Mayer, R. E. (2009). *Multimedia learning* (2d ed.). New York: Cambridge University Press.
- McLaren, B. M., DeLeeuw, K. E., & Mayer, R. E. (2011). A politeness effect in learning with web-based intelligent tutors. *International Journal of Human-Computer Studies*, 69(1–2), 70–79.
- McNamara, D. S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, 34(2), 147–171.

- 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: Springer.
- Moreno, R., & Mayer, R. E (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, 97(1), 117–128.
- Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review*, 19(3), 309–326.
- National Research Council (2010). The role of simulations and games in science assessment. In M. A. Honey & M. Hilton (Eds.), *Learning science through computer games and simulations* (pp. 87–104). Washington, DC: National Academies Press.
- Olney, A., D’Mello, S., Person, N., Cade, W., Hayes, P., Williams, C., Lehman, B., & Graesser, A.C. (2012). Guru: A computer tutor that models expert human tutors. In S. A. Cerri & B. Clancey (Eds.), *Proceedings of Intelligent Tutoring Systems (ITS) 2012* (pp. 256–261). Berlin: Springer.
- O’Neil, H. F., & Perez, R. S. (2008). *Computer games and team and individual learning*. Amsterdam: Elsevier.
- Pardos, Z. A., Dailey, M. D., & Heffernan, N. T. (2011). Learning what works in ITS from non-traditional randomized controlled trial data. *International Journal of Artificial Intelligence in Education*, 21(1), 47–63.
- Razzaq, L., & Heffernan, N. T. (2010). Hints: Is it better to give or wait to be asked? In V. Alevin, J. Kay, & J. Mostow (Eds.) *Proceedings of Intelligent Tutoring Systems (ITS) 2010* (pp. 349–358). Berlin: Springer.
- Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14(2), 249–255.
- Roll, I., Alevin, 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.
- 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(1–2), 115–133.
- Schatz, S., Oakes, C., Folsom-Kovarik, J. T., & Dolletski-Lazar, R. (2012). ITS + SBT: A review of operational situated tutors. *Military Psychology*, 24(2, SI), 166–193.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and instruction*, 13(2), 141–156.
- Shih, C. (2012). Zero tolerance cue angle analysis and its effect on successive sink rate of a low cost billiard reposition control tutoring system. *Knowledge-Based Systems*, 30, 17–34.
- Shute, V.J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189.
- Silverman, B. G., Pietrocola, D., Nye, B. D., 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.

- Singh, R., et al. (2011). Feedback during web-based homework: The role of hints. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of Artificial Intelligence in Education (AIED) 2011* (pp. 328–336). Berlin: Springer.
- 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*.
- Stottler, R., & Panichas, S. (2006). A new generation of tactical action officer intelligent tutoring system (ITS). In *Proceedings of Interservice/Industry Training, Simulation and Education Conference (IITSEC) 2006*. Orlando, FL.
- Suraweera, P., & Mitrovic, A. (2004). An intelligent tutoring system for entity relationship modelling. *International Journal of Artificial Intelligence in Education*, 14(3–4), 375–417.
- U. S. Department of Education, Institute of Education Sciences (2009). Cognitive Tutor Algebra 1. What Works Clearinghouse. Retrieved from ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_cogtutor_072809.pdf.
- 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.
- VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31, 3–62.
- VanLehn, K., et al. (2005). The Andes physics tutoring system: Five years of evaluations. In G. I. McCalla and C.-K. Looi (Eds.), *Proceedings of the Artificial Intelligence in Education (AIED) 2005* (pp. 678–685). Amsterdam: IOS Press.
- Wang, Y. H., Young, S. S. C., & Jang, J.-S. (2009). Evaluation of tangible learning companion/robot for English language learning. In *Proceedings of the International Conference on Advanced Learning Technologies (ICALT) 2009* (pp. 322–326). Riga, Latvia: IEEE Press.
- Ward, W., Cole, R., Bolaños, D., Buchenroth-Martin, C., Svirsky, E., Van Vuuren, S., Weston, T., Zheng, J., & Becker, L. (2011). My Science Tutor: A conversational multimedia virtual tutor for elementary school science. *ACM Transactions of Speech and Language Processing*, 13, 4–16.
- Woolf, B. P. (2009). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Burlington, MA: Elsevier.