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.*

  *Tut*or: *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 feed-back) 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).

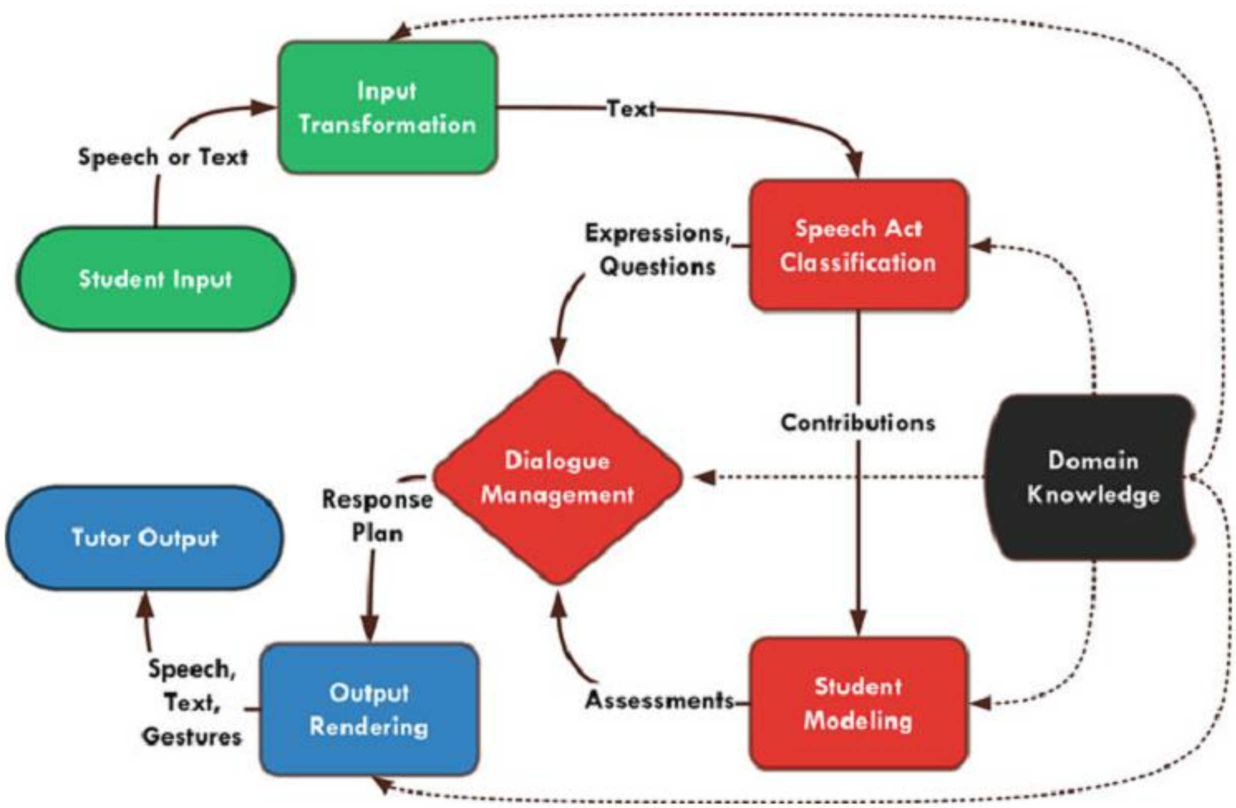

Figure 1. Interactions of various portions of dialogue-based ITS (D'Mello & Graesser, 2013).

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

We would like to acknowledge the contributions of the U.S. Army Research Laboratory (ARL) in conjunction with the Learning in Intelligent Tutoring Environments (LITE) Laboratory and the GIFT project. The research was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428, REESE 0633918, ALT-0834847, DRK-12-0918409, 1108845), the Institute of Education Sciences (R305H050169, R305B070349, R305A080589, R305A080594, R305A090528, R305A100875, R305C120001), ARL, and the Office of Naval Research. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF, Institute of Education Sciences (IES), or the Department of Defense (DoD).

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.