7-28-2010

# The Effects of Cognitive Disequilibrium on Student Question Generation While Interacting with AutoTutor

Jeremiah R. Sullins

To the University Council:

The Dissertation Committee for Jeremiah Sullins certifies that this is the final approved version of the following electronic dissertation: "The Effects of Cognitive Disequilibrium on Question Generation while Interacting with AutoTutor."

Art Graesser, Ph.D.
Major Professor

We have read this dissertation and recommend its acceptance:

Danielle McNamara, Ph.D.

Scotty Craig, Ph.D.

Deborah Lowther, Ph.D.

Accepted for the Graduate Council:

Karen D. Weddle-West, Ph.D.
Vice Provost for Graduate Programs

# THE EFFECTS OF COGNITIVE DISEQUILIBRIUM ON STUDENT GENERATED QUESTIONS WHILE INTERACTING WITH AUTOTUTOR

by

Jeremiah Ray Sullins

A dissertation

Submitted in Partial Fulfillment Presented of the

Requirements for the Degree of

Doctor of Philosophy

Major: Experimental Psychology

The University of Memphis

August 2010

**ABSTRACT**

Sullins, Jeremiah Ray. The University of Memphis. August, 2010. The Effects of Cognitive Disequilibrium on Question Generation While Interacting with AutoTutor. Major Professor: Art Graesser, Ph.D.

The purpose of this study was to test the effects of cognitive disequilibrium on student question generation while interacting with an intelligent tutoring system. Students were placed in a state of cognitive disequilibrium while they interacted with AutoTutor on topics of computer literacy. The students were tutored on three topics in computer literacy: hardware, operating system, and the internet. During the course of the study a confederate was present to answer any questions that the participant may have had. Additional analyses examined any potential influence the confederates had on student question asking. Lastly, the study explored the relationship between emotions and cognitive disequilibrium. More specifically, the study examined the temporal relationship between confusion and student generated questions. Based on previous cognitive disequilibrium literature, it was predicted that students who were placed in a state of cognitive disequilibrium would generate a significantly higher proportion of question than participants who were not placed in a state of cognitive disequilibrium. Additionally, it was predicted that students who were placed in a state of cognitive disequilibrium would generate “better” questions than participants who were not in a state of cognitive disequilibrium. Results revealed that participants who were not placed in a state of cognitive disequilibrium generated a significantly higher proportion of questions. Furthermore, there were no significant differences found between participants for deep or intermediate questions. Results did reveal significant main effects as a function of time

for certain action units. Lastly, it was discovered that certain measures of individual differences were significant predictors of student question generation.

TABLE OF CONTENTS

# LIST OF TABLES

# INTRODUCTION

Question generation has received a great deal of attention in recent years from researchers in the fields of computer science (Heilman & Smith, 2010), psychology (Graesser, Ozuru, & Sullins, 2009; Rus & Graesser, 2009; Sullins & McNamara 2009) and education. Question generation is believed to play a crucial role in a variety of cognitive faculties, including comprehension (Collins, Brown, & Larkin, 1980; Graesser, Singer, & Trabasso, 1994) and reasoning (Graesser, Baggett, & Williams, 1996; Sternberg, 1987). Asking good questions has been shown to lead to improved memory and comprehension of material in school children and adult populations (Rosenshine, Meister, & Chapman, 1996). Available research suggests that learning how to ask good questions should be taught at an early age but all ages benefit from question generation training (Wisher & Graesser, 2007).

Sadly, it is well documented that the ideal scenario of a curious question asker does not match reality. Students are unspectacular at monitoring their own knowledge deficits and their question generation is both infrequent and unsophisticated (Baker, 1979; Dillon, 1988; Graesser & Person, 1994; Van der Meij, 1988). Graesser and Person (1994) reported that an individual student asks approximately 1 question in 7 hours of class time (around 1 question per day).  Most of these questions are not good questions, so the quality is also disappointing.

However, imagine a scenario in which a student is given a learning task in which obstacles are in place to challenge the student. During the learning task the student may encounter concepts that are unfamiliar to them or ideas that are contradictory. This unfamiliar or contradictory content may place the student in a state of *cognitive* *disequilibrium*. Cognitive disequilibrium is a situation in which learners are presented with obstacles to goals, anomalous events, contradictions, discrepancies, and/or obvious gaps in knowledge within the learner's range of knowledge (Graesser, Lu, Olde, Cooper-Pye, & Whitten, 2005; Otero & Graesser, 2001). Perhaps the most salient example of cognitive disequilibrium is when the student is forced to handle an obstacle that prevents them from achieving a goal. This disparity between student knowledge and outside events/information is expected to frequently create a state of confusion and requires extra processing in order to resolve the discrepancy. Students who are in a state of cognitive disequilibrium are potentially in a perfect scenario to significantly increase the amount of questions asked in order to restore cognitive equilibrium.

Emotions have also been shown to be associated with goal blockage and cognitive disequilibrium (Baker, D'Mello, Rodrigo, & Graesser, 2010; D'Mello & Graesser, 2010; Graesser, D'Mello, Chipman, King, & McDaniel, 2007; Graesser et al., 2006). These studies have tracked a wide variety of emotions during students' interactions with intelligent tutoring systems and problem solving games. The emotions they experience depend on the learning challenges, the amount of changes they experience, the blockage of important goals, and other factors. Negative emotions, such as confusion and frustration, are likely to occur when learners confront such events as anomalies, obstacles to goals, or other stimuli that fail to match expectations (Baker et al., 2010; D'Mello et al., 2010; Graesser et al., 2007; Graesser, Jackson, & McDaniel, 2007; Graesser et al., 2006). Some central hypotheses that will be explored in this dissertation are that the emotion of confusion is associated with cognitive disequilibrium and can be detected and measured by observing facial expressions. In summary, there is an important alignment between cognitive disequilibrium, emotions, and question asking.

The primary focus of this dissertation is on testing a systematic relationship between cognitive disequilibrium (which is both a cognitive and affective state) and student question generation. An experiment was conducted that manipulated cognitive disequilibrium and assessed the impact on the quantity and quality of generated questions. A secondary question was whether the disequilibrium was manifested in the facial expressions that reflect confusion.

This dissertation begins with a description of question asking and its relationship with cognitive disequilibrium. It then discusses the growing area of emotion research in the field of cognitive science, with a particular emphasis on the affective state of confusion that results from cognitive disequilibrium. In order to test the hypotheses, an experiment was conducted that manipulated different scenarios of cognitive disequilibrium in order to determine its effect on confusion and student generated questions.

*Question Asking and Cognitive Disequilibrium*

It is well documented that students and adults have trouble generating questions (Dillon, 1988; Graesser & Person, 1994; Wisher & Graesser, 2005). Of the questions that are generated, the majority are shallow questions rather than questions that require deep reasoning. A deep reasoning question is one which integrates content and that fosters understanding of the components and mechanisms being covered (Craig, Vanlehn, & Chi, in press). Deep reasoning questions are questions that typically invite lengthier answers (usually around a paragraph in length) and often start with words such as *why*, *how*, or *what-if* (Graesser et al., 2010). These questions are aligned with the higher levels of Bloom's taxonomy (1956) and the long-answer question categories in the question taxonomy proposed by Graesser and Person (1994).

In order to illustrate the difference between shallow reasoning questions and deep reasoning questions, consider an example of each. An example of a shallow reasoning question, according to Graesser and Person (1994), would be "Does the CPU use RAM when running an application?". The reason for categorizing this type of question as "shallow" is because it does not require substantial thought on the student's part; indeed, the student could answer it by simply guessing yes or no. In contrast, a deep reasoning question would be "How does the CPU use RAM when running an application?". The reason for categorizing this question as "deep" is because the student must use the knowledge known about computers to articulate the causal mechanisms that relate two components in the operating system. They not only need to generate a nontrivial amount of content, but must be able to reason about complex causal mechanisms.

Graesser and Person (1994) estimated that a typical student asks only .11 questions per hour in a traditional setting, such as a classroom. There are several possible explanations as to why students do not ask many questions. These include the lack of prior domain knowledge, high social editing, and insufficient training/modeling. The first explanation for the lack of student questions might be due to insufficient prior knowledge so they are incapable of monitoring the fidelity of knowledge. For example, Miyaki and Norman (1979) posit that students need a large amount of knowledge to detect when they do not understand something. Because of this, students simply do not know that they do not understand and therefore do not ask questions. The second possible explanation for a low amount of student questioning is due to social editing. Students may not ask questions because they are afraid of looking ignorant in front of their peers and losing social status. The third reason for a low number of student questions has to do with the training they receive. Graesser and Person (1994) point out that 96% of questions that occur in the classroom come from the teacher and most of the questions are shallow. Therefore, students in a typical classroom are not provided with examples of good deep-reasoning questions from the teachers. And of course, given the above statistics on student question asking, students rarely observe other students asking questions. Good student role models are essentially absent.

There is little doubt that students have difficulty asking questions on their own accord, but there is reason to be optimistic. In particular, the research on relationships between questions and cognitive disequilibrium indicate that under the correct circumstances the rate of student questioning can significantly increase or improve in quality (Graesser & McMahen, 1993; Otero & Graesser, 2001; Wisher & Graesser, 2007). As mentioned previously, cognitive disequilibrium is a situation in which learners encounter obstacles to goals, which places them in a state of inquiry until cognitive equilibrium is restored.

Graesser and Olde (2003) investigated the questions asked by college students when an everyday device breaks down. The broken devices were selected to induce cognitive disequilibrium. Participants were asked to read six illustrated texts taken from David Macaulay's *The Way Things Work* (1998). Each text was read for five minutes and then participants encountered a scenario in which the device breaks down (e.g., "the key turns but the bolt does not move"). Graesser and Olde correlated the quality of student questions with an objective device comprehension questionnaire. Correlational analysis revealed that the most robust predictor of students' comprehension was not the quantity of student questions but rather the quality of the student questions. The students who asked deeper questions (i.e., questions that explained the causes of the breakdowns) ended up having a better understanding of how the devices functioned. These results are consistent with other research that suggests that simply asking a large number of questions does not necessarily correlate with understanding of the material (Fishbein, Eckart, Lauver, Van Leeuwen, & Langmeyer, 1990).

A study conducted by Graesser, Lu, Olde et al. (2005) explored the effects of cognitive disequilibrium and student generated questions as a function of time. Participants' eye movements were tracked while they read illustrated texts about everyday devices. Participants were then placed in a state of cognitive disequilibrium by reading about a breakdown scenario of the device. If cognitive disequilibrium does in fact trigger questions, then participants should have fixated (based on eye tracking) on the faulty parts of the device prior to or during a question. Results revealed that good comprehenders did fixate significantly more on the faulty locations both *before* and *during* question generation, but not immediately after a question was asked. Graesser and McMahen (1993) explored question asking while participants read expository texts with varying degrees of anomalous information. In Experiment 1, upper division undergraduate students solved quantitative problems. The participants received three versions of the quantitative problems: original, deleted, and irrelevant. The *original* version was the exact copy of the problem out of a textbook (e.g., "The Bears won 40 games and lost 24, while the Bulls won 32 games and lost 18. Which team had the higher percentage of wins?"). The *deletion* version was exactly the same as the original, except that a phrase, clause, or sentence was removed (e.g., The Bears won 40 games and lost 24, while the Bulls won 32. Which team had the higher percentage of wins?). The *subtle irrelevancy* version was the same as the original problem except that an irrelevant clause or sentence was added (e.g., The Bears won 40 games and lost 24, while the Bulls won 32 games and lost 18. Nearly 20% of the games almost resulted in ties. Which team had the higher percentage of wins?). Results revealed no significant difference between the three conditions with respect to the number of generated questions. However, further analyses explored transformation questions (i.e., questions that addressed the changes in the text). The mean number of transformation-relevant questions per problem was significantly higher in the deletion condition than in the subtle irrelevancy condition. The results from this study suggest that having students read text with anomalous information that presented an obstacle may have placed the students in cognitive disequilibrium, which in turn significantly increased higher quality questions.

Experiment 2 of Graesser and McMahen (1993) required subjects to generate questions and solve quantitative problems that were presented in one of five versions: original, deletion, contradiction, salient irrelevancy, and subtle irrelevancy. The deletion and subtle irrelevancy versions were constructed in the same manner as Experiment 1. The *contradiction* version contained a statement or added a statement to the original version that directly contradicted previous statements (e.g., The Bears won 40 games and lost 24, while the Bulls won 32 games and lost 18. The Bulls won more games than the Bears. Which team had the higher percentage of wins?). The *salient irrelevancy* version added a statement that was both irrelevant in solving the problem and to the context of the problem (e.g., The Bears won 40 games and lost 24, while the Bulls won 32 games and lost 18. Several Bear team members were always getting into arguments with some members of the Bulls team. Which team had the higher percentage of wins?). Results revealed significant differences in the amount of questions asked among the conditions. Post-hoc tests indicated that the deletion and salient irrelevancy conditions significantly differed from the original, whereas the contradiction and subtle irrelevancy conditions did not. Furthermore, as in Experiment 1, the number of transformation-relevant questions was examined. Results revealed that the deletion and contradiction conditions produced more transformation relevant questions than did the two irrelevancy conditions.

These three studies support the claim that anomalous information may place students in a state of cognitive disequilibrium (confusion) and significantly increase the proportion and/or quality of questions asked. These results provide a framework for the current research which will test the hypothesis that cognitive disequilibrium can influence question generation. More specifically, these results demonstrate that leaving out critical information, presenting contradictions, or adding puzzling information is an effective method for manipulating cognitive disequilibrium and influencing the resulting questions.

*AutoTutor*

A tutoring system called AutoTutor was used in this experiment to serve as a learning environment for investigating cognitive disequilibrium, question generation, and emotions. AutoTutor is a fully automated tutor that holds conversations with learners in natural language and that simulates the dialogue moves of human tutors (Graesser, Chipman, Olney, & Haynes, 2005). AutoTutor presents students with a series of challenging problems (or main questions), each requiring approximately 3 to 7 sentences of information for a correct answer. When presented with a problem, students typically respond with answers that are only one word to two sentences in length. In order to guide students in their construction of an improved answer, AutoTutor actively monitors students' knowledge states and engages learners in a turn-based dialogue. AutoTutor adaptively manages the tutorial dialogue by providing feedback, pumping the learner for more information, giving hints, correcting misconceptions, answering questions, and summarizing answers. Table 1 presents a list of questions, expectations, misconceptions, and dialogue moves of AutoTutor, along with definitions. As students move through their learning sessions, AutoTutor adheres to constructivist theories of learning by guiding students to actively construct answers to difficult questions, as opposed to simply presenting the solution to the students.

In order to correctly answer a main question, students typically have to construct dozens of dialogue turns. After each student dialogue turn, AutoTutor compares the learner's response to expectations and common misconceptions for that main question. AutoTutor leads the student through the relevant expectations and assists their learning by using a specific dialogue structure referred to as a hint-prompt-assertion cycle. This cycle is designed to gradually shift the cognitive burden from student to the tutor. Specifically, AutoTutor first presents the student with a hint that requires the student to provide the majority of the information (5 in Table 1). If the student does not supply an adequate answer, AutoTutor then prompts the student for more information (7 in Table 1). If the student is still unable to provide the correct answer desired by the prompt, AutoTutor provides an assertion (9 in Table 1). If the student provides a satisfactory answer at any point during the hint-prompt-assertion cycle, AutoTutor exits the cycle and move to the next expectation (see below). If the student provides a satisfactory answer that covers the last remaining expectation, then AutoTutor asks the student to restate a complete answer and subsequently provides a summary of what an ideal answer would look like (11 in Table 1). For a detailed discussion of AutoTutor's dialogue mechanisms and strategies, see Graesser et al. (2005).

It is important to emphasize that AutoTutor is not merely a scripted information delivery system, but rather is very adaptive to the student's knowledge and the dialogue history. AutoTutor provides the student with feedback (positive, neutral positive, neutral, neutral negative, and negative) after the student expresses information in a conversational turn. AutoTutor selects hints, prompts, and assertions in a dynamic fashion that is sensitive to the responses of the student.

Table 1

*Tutorial Dialogue Moves and Assessments within AutoTutor*

| Function | # | Dialogue move category | Description of dialogue moves |
|---|---|---|---|
| Questions & Expected Answers | 1 | Main Tutoring Questions | Provides an example scenario and asks the student to provide an answer and an explanation |
| | 2 | Expectations | A set of several sentence-like statements that address any important and required information necessary to completely answer a specific main tutoring question. |
| | 3 | Misconceptions | A set of sentence-like statements that include information related to a possible conceptual misunderstanding of the content addressed by a main tutoring question. |
| Dialogue Moves | 4 | Pumps | Information request with minimal information provided. e.g. "What else" |
| | 5 | Hints | Provides a guided suggestion to the student to fill in a proposition |
| | 6 | Hint completions | Contains the ideal answer to a given hint |

(table continues)

Table 1 (continued)

| Function | # | Dialogue move category | Description of dialogue moves |
|---|---|---|---|
| | 7 | Prompts | Requests the student to fill in a missing content word |
| | 8 | Prompt completions | Contains the target word for specific prompts |
| | 9 | Assertions | Asserts information about an expectation |
| | 10 | Corrections | Corrects the student's misconception |
| | 11 | Summaries | Provides a summary of the ideal answer to a main tutoring question |
| Immediate Short Feedback | 12 | Positive | Provides terms such as: "good job", "correct" |
| | 13 | Neutral Positive | Provides terms such as: "yeah", "hmm right" |
| | 14 | Neutral | Provides terms such as: "uh huh", "alright" |
| | 15 | Neutral Negative | Provides terms such as: "possibly", "kind of" |
| | 16 | Negative | Provides terms such as: "not quite", "no" |

*Previous Findings*

AutoTutor has been evaluated on numerous studies evaluating learning gains (Graesser, Lu, Jackson, Mitchell, Ventura, Olney, & Louwerse, 2004; Jackson, Mathews, Lin & Graesser, 2003; Van Lehn et al., 2007). The studies evaluating AutoTutor have continually shown robust learning for students who interact with the tutor. AutoTutor has been compared to various controls: no tutoring, reading of relevant textbook chapters, reading of ideal custom text, other intelligent tutoring systems, and expert human tutors. The previous studies have revealed encouraging results regarding student learning. On average, the studies have resulted in positive effect sizes for learning ($d$=.80) for conceptual physics and computer literacy, when compared to pretest scores and the reading of a textbook for an equivalent amount of time (Graesser et al., 2004, VanLehn et al., 2007).

A set of structural features are present in the AutoTutor's interface in all conditions. These include: (a) a main question pane, for displaying the text of the current question or problem that students were attempting to answer (see top of Figure 1); (b) an animated pedagogical agent for communicating AutoTutor's dialogue moves to the learner; (c) an image pane, which displays images or diagrams to help students answer the current problem (not all problems contain an image or diagram) and (d) a dialog history box where students can observe what AutoTutor has previously said and what the student has typed. The animated agent speaks to participants with a computer-generated voice and gives nonverbal feedback through realistic facial gestures. In all conditions, AutoTutor communicates its dialogue moves orally. At all times, students are able to read the text of tutorial dialogue history.

The current study used AutoTutor as the learning environment and the research goal was to put the student in cognitive disequilibrium. Cognitive disequilibrium was induced by having AutoTutor deliver two types of incorrect information. First, AutoTutor presented incorrect short feedback after student turns. Second, AutoTutor presented incorrect information as answers to question prompts. More will be said about these manipulations in the Methods section.

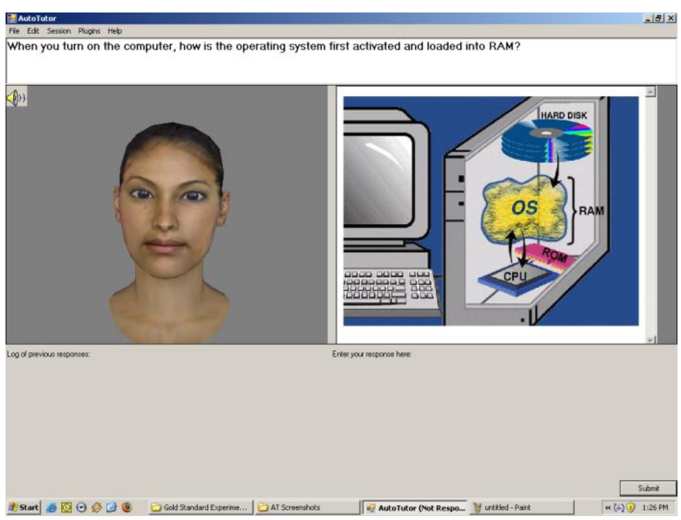


*Figure 1*. Example screenshot of AutoTutor for computer literacy.

*Confusion and Learning*

It is often assumed that emotions are merely a motivational factor and have no direct influence on complex learning. However, in recent years the link between emotions and learning has received a growing interest in education, psychology, computational linguistics, and artificial intelligence (Breazeal, 2003; Conati, 2002; Craig, Graesser, Sullins, & Gholson, 2004; De Vincente & Pain, 2002; Kort, Reilly, & Picard, 2001; Lepper & Woolverton, 2002; Litman & Forbes-Riley, 2004; Picard 1997; Wang et al., 2005).

One seminal contribution to the field of emotion detection was the work conducted by Ekman and Friesen (1978). Ekman and Friesen developed the facial action coding system (FACS). The purpose of the FACS was to identify "basic emotions" located on the face and to associate these emotions with the muscles that produce them. The basic emotions identified by the FACS are happiness, sadness, surprise, disgust, anger and fear. Each muscle movement on the face is referred to as an action unit. Ekman and Friesen identified approximately 58 action units in total. However, all of the emotions studied were used with static pictures and did not include changing expressions over time. Moreover, researchers interested in exploring the relation between cognition and emotions have concluded that these "basic emotions" do not occur frequently during complex learning (Baker et al., 2010; D'Mello & Graesser, 2010; Graesser et al., 2006). Researchers have expanded upon the basic set of emotions identified by Ekman, and have begun to conduct studies that empirically test the link between emotions and learning.

A study that directly tested the relationship between emotions and learning was conducted by Graesser, D'Mello, Craig et al. (2008). In this study five participants interacted with AutoTutor for 90 minutes. The students were asked to emote-aloud after they were trained on the relevant emotions, which consisted of anger, boredom, confusion, contempt, curiosity, disgust, eureka, and frustration. Results revealed a small number of occurrences for anger (n = 17), contempt (n = 8), curiosity (n = 1), and disgust (n = 5) so these emotions were removed from subsequent analyses. In contrast, sufficient frequencies of emotions were available for boredom (n = 40), confusion (n = 53), delight (n = 28), and frustration (n = 49). Analyses revealed significant correlations between AutoTutor's dialog (i.e., session information, quality of learner contributions, directness of AutoTutor supplying information and AutoTutor feedback) and three emotions: confusion, eureka, and frustration.

Additionally, a study conducted by Craig et al. (2004) directly explored the relation between confusion and learning. In this study, the authors were interested in examining the following emotions: frustration, boredom, flow, confusion, eureka, and neutral. During the study, participants interacted with AutoTutor on the topic on computer literacy with a particular focus on hardware. Both before and after their interaction, participants completed a multiple choice test on computer literacy in order to determine any possible learning gains. During participants' interactions with AutoTutor, a rater would code which of the six emotions they felt the participants was experiencing (5 minute intervals). A total of 5 raters were used during the study but only one rater was present at a time during each individual learning session. Results revealed that learning gains were positively correlated with confusion and flow, but negatively correlated with frustration.

These results suggest that not only are emotions present during learning but that emotions are linked to cognition. More specifically, these results provide evidence that emotions (i.e., confusion, frustration and boredom) are a crucial component during learning. When students experience confusion, they are placed in a state of inquiry which in turn drives questions (Graesser, Lu, Olde et al., 2005). Furthermore, these studies suggest cognitive disequilibrium may be manifested by confusion. When students are encountering information that they may not know or contradicts what they already know it is placing them in a state of cognitive disequilibrium which is manifested on the face through the display of confusion.

*Emotions and Cognitive Disequilibrium*

Research has revealed that confusion is a dominant factor when studying learner emotions while interacting with intelligent tutoring systems. In numerous studies, a significant positive correlation has been revealed between learning and confusion (Craig et al., 2004; Graesser, D'Mello, Chipman et al., 2007; Graesser, D'Mello, & Craig, 2008). This relation between confusion and learning provides additional support for the importance of cognitive disequilibrium (Graesser & Olde, 2003; Otero & Graesser, 2001). During students' interactions with an intelligent tutoring system, they may encounter new incoming information which reveals knowledge gaps or discrepancies with existing prior knowledge. Such obstacles are expected to place the student in a state of cognitive disequilibrium, which in turn triggers confusion, and also the inquiry that motivates asking questions until equilibrium is restored. The relative time-course of confusion and inquiry is uncertain.

In one example, Craig, D'Mello, Witherspoon, and Graesser (2008) investigated two methods of identifying affective states while students interacted with AutoTutor. During their interaction with AutoTutor, learners were told to emote-aloud by expressing the affective state that they were feeling while they were interacting with the tutoring system. Two trained judges used the FACS to independently rate the action units that occurred 3 seconds prior to each learner utterance and during the utterance. Analyses were able to determine significant relationships with action units for frustration, boredom, and confusion. More specifically, it appeared that action units 1 (inner brow raiser), 2 (outer brow raiser), and 14 (dimpler) were associated with boredom. Confusion displayed action units 4 (lowered brow) and 7 (tightened lids). Boredom displayed a significant association with action unit 43 (eye closure). These results provide evidence that specific facial action units do occur with particular emotions.

A study conducted by McDaniel et al. (2007) used different methodologies to explore emotions while students interacted with AutoTutor. They investigated facial features that reflect the emotions that accompany deep-level learning of conceptual material. Participants interacted with AutoTutor for 32 minutes while the system recorded the participant's face, body posture, and the computer screen. The participant's session was then observed and scored by four different judges (the learner, a peer, and two trained judges). Following the previous research on emotions related to learning, the emotions assessed in this study were boredom, confusion, flow, frustration, delight, neutral, and surprise. Raters watched the learners' sessions and were told to rate any emotion they saw every 20 seconds (mandatory judgments). Raters were also instructed to rate any emotions they observed within the 20 second interval (voluntary judgments).

Results revealed that confusion was manifested by action unit 4 (lowering of the brow) and action unit 7 (tightening of the eyelids).

The experiment in this dissertation expanded on the results found in the cognitive disequilibrium literature. First, we manipulated different scenarios in which students were placed in a state of cognitive disequilibrium and tested the hypothesis that there would be significant increases in the amount of questions generated. Second, we assumed that the questions would be influenced by a social aspect, namely by the presence of a confederate during the learning session who answered questions. Third, the dissertation explored the temporal relationships between question generation and the emotion/affective state of confusion (cognitive disequilibrium), as manifested by facial action units. Do students ask a question at the exact same time they experience confusion? Or does confusion precede questions? Do students sometimes experience confusion, but do not ask questions because of the social presence of the other person and their resistance to appear unknowledgeable?

## EXPERIMENT: COGNITIVE DISEQUILIBRIUM AND QUESTION ASKING

This study was designed to explore question generation when students were placed in a state of cognitive disequilibrium while interacting with AutoTutor. Since AutoTutor is incapable of answering any question that a student may ask, the student was accompanied by a human who answered questions. This confederate was a learning companion who assisted the student in his or her learning. The social presence of the learning companion would undoubtedly influence the frequency and quality of questions asked by participants. However, this presence of a learning companion was held constant in all phases of the experiment, as opposed to the induction of cognitive disequilibrium which was experimentally manipulated. The confederate in this study was an undergraduate student with prior knowledge and training of the AutoTutor content. Lastly, the dissertation investigated the temporal relation between confusion (i.e., action units 4 & 7) and question asking.

The version of AutoTutor used in the current study was designed to tutor three topics in computer literacy (hardware, operating systems, and internet). Each topic consisted of four main questions. Example main questions include, "How is the packet switching model of message transmission like the postal system?" from the internet topic and "How would you design an operating system that can manage memory demands from multiple concurrent jobs?" from the operating systems topic. Participants were tutored with one of two versions of AutoTutor on all three available computer literacy topics. In a *cognitive disequilibrium AutoTutor* condition, students interacted with a version of AutoTutor that sometimes provided incorrect feedback in response to student contributions and also incorrect information. For example, in response to a student's correct answer, instead of AutoTutor responding with short positive feedback (e.g., "Good job!"), AutoTutor would respond with short negative feedback ("Not quite."). Regarding the incorrect information, AutoTutor would assert a statement that is incorrect (e.g., "a digital camera is never used for uploading photos to the Web"). In a second condition called *regular AutoTutor*, students were tutored while interacting with a standard version of AutoTutor, which delivered correct feedback and information and which served as a control.

A second research goal of this dissertation was to test the hypothesis that the cognitive-affective state of confusion (which is considered to be a manifestation of cognitive disequilibrium) directly triggers or is otherwise associated with student questions. Research has confirmed that particular facial action units are diagnostic of confusion (namely facial action units 4 and 7), so the study could record the moment when a student encounters cognitive disequilibrium. All interactions with both AutoTutor and the confederate were recorded using a standard webcam and desktop microphone. Following the completion of the study, each individual session was coded using the facial action coding system in order to explore answers to two questions: (1) Does presenting students with contradictory information and false feedback place students in cognitive disequilibrium? and (2) Do students show a different proportion of confusion indicators (i.e., action units 4 and 7) before asking a question, during the question, versus after asking a question. Answers to these questions were expected to shed additional light on the relative timing of confusion and question asking.

Lastly, it is plausible to expect that individual differences such as motivation and personality are robust predictors of learning in the context of particular learning environments. In order to determine if such individual differences had any influence on the quantity or quality of questions, we explored the correlations between four tests of individual differences and questions asking.

*Experiment Predictions*

Previous research on cognitive disequilibrium provides some foundation to offer predictions on the results of this experiment. One prediction is that college students will generate more questions in the *cognitive disequilibrium AutoTutor* condition than in the *regular AutoTutor* condition. According to Graesser and McMahen (1993), presenting some types of anomalous information to learners can lead to an increase in student generated questions. This prediction specifies that the mean frequency of questions in the *cognitive disequilibrium AutoTutor* condition should be greater than the mean frequency in the *regular AutoTutor* condition: cognitive disequilibrium > control.

Additionally, predictions can be made based on the quality of questions participants ask. After inducing cognitive disequilibrium in participants, Graesser and Olde (2003) discovered that the most robust predictor of students' comprehension was not the quantity of student questions but rather the quality of the student questions. In other words, the students who asked deeper questions (i.e., questions that explained the causes of the breakdowns) ended up having a better understanding of how the devices functioned. This prediction specifies that the mean frequency of high quality questions in the *cognitive disequilibrium AutoTutor* condition should be greater than the mean frequency of high quality questions in the *regular AutoTutor* condition.

Some predictions can be made regarding the relative timing of confusion and question asking. There is one study that motivates some predictions about cognitive disequilibrium and question generation as a function of time (Graesser, Lu, Olde et al. 2005). According to Graesser et al. (2005), if cognitive disequilibrium triggers questions, then at the moment when contradictory information is presented to the student, an increase in confusion indicators (i.e., action units 4 and 7) should occur both *before* and *during* a student generated question. However, confusion should not occur as frequently *after* a student question.

Nevertheless, there is an alternative plausible scenario in which confusion indicators could increase *after* a question. This may occur when a student asks a clarification question that is not considered a genuine information seeking question. The clarification question may unveil a misconception to the student that the student was unaware of until the question was asked. The misconception essentially vaults the student into a state of cognitive disequilibrium that yields increased confusion. An analysis of the timing of confusion and questions should shed light on the relative timing of the different theoretical components.

*Methods*

*Participants*. Participants consisted of 48 undergraduate students. The participants were recruited from the Psychology Subject Pool at the University of Memphis and received course credit in return for their participation.

*Materials*. The test materials consisted of two tests of computer literacy and a post-interaction questionnaire. Participants were tested on their knowledge of computer literacy topics both before and after the tutorial session (pretest and posttest, respectively). Based on these test scores, learning gains were computed to determine the amount of knowledge that students acquired as a result of the tutorial session. The testing materials were adapted from computer literacy tests used in previous experiments involving AutoTutor, and were comprised of questions that assessed students' knowledge of all three computer literacy topics. Each test contained 18 multiple-choice questions: 6 questions on hardware, 6 questions on operating systems, and 6 questions on Internet. Participants completed alternate test versions for pretest and posttest. The two test versions, composed of different questions, tested learners on the same subject matter and content. The assignment of test versions to pretest versus posttest was counterbalanced across participants.

The knowledge tests were designed to tap into deep knowledge through a 4-alternative multiple choice format. These deep-level multiple-choice questions were designed to assess students on causal reasoning, inferences, and the underlying functional organization of concepts. Questions that assess deep levels of knowledge can be contrasted with those that assess shallow levels of knowledge by simply asking students to recall previously presented information, definitions, and facts. A complete list of items of the multiple-choice questions used in the testing materials can be seen in Appendix A.

Research has suggested that motivation and personality are robust predictors of learning in the context of particular learning environments, so two tests of individual differences were used for the study. The first test of individual differences was the motivated strategies for learning questionnaire (MSLQ) that was designed to measure motivation (Pintrich, Smith, Garcia, & McKeachie, 2001). Table 2 shows some example questions on this test. The MSLQ is used for assessing college students' motivational orientations and their different learning strategies during a college course. In other words, the two main constructs being measured by the MSLQ are motivation and learning strategies. The MSLQ contains two different sections and consists of 81 items (31 questions addressing motivation and 50 items addressing learning strategies). Participants answer each question using a 7-point scale where 1 represents "not at all true of me" and 7 represents "very true of me". More specifically, the MSLQ measures participants on the following: 1) Intrinsic Goal Orientation which refers to the students' perception of the reasons why she is engaged in a learning task; 2) Extrinsic Goal Orientation which complements Intrinsic Goal Orientation and concerns the degree to which the student perceives herself to be participating in a task for reasons such as grades, rewards, performance, and competition; 3) Task Value which refers to the students' evaluation of how interesting, how important, and how useful the task is; 4) Control of Learning Beliefs which refers to the students beliefs that their efforts to learn will result in positive outcomes; 5) Self Efficacy for Learning and Performance which refers to performance expectations; 6) Test Anxiety which refers to a student's negative thoughts that disrupt performance; 7) Rehearsal which is a strategy used involving reciting items from a list to be learned; 8) Elaboration which is a strategy that includes paraphrasing, summarizing, creating analogies, and generative note taking; 9) Organization which helps the learner select appropriate information and also construct connections among the information to be learned; 10) Critical Thinking which refers to the degree to which students report applying previous knowledge to new situations in order to solve the problem; 11) Metacognitive Self Regulation which refers to the awareness knowledge and control of cognition; 12) Time and Study Environment which involves scheduling, planning, and managing one's study time; 13) Effort Regulation which includes students' ability to control their effort and attention in the face of distractions; 14) Peer Learning which includes dialogues with peers; and 15) Help Seeking which involves the support of others.

A second test measuring individual differences was the Big Five personality test, which was used to determine whether any interactions existed between learners' personality and questions asked (see Table 2 for example questions). The Big Five personality test was designed to measure the following personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism. The Big Five personality test consisted of 45 items which were rated using a 5-point scale with 1 representing “strongly disagree” and 5 representing “strongly agree” (see Appendix B for complete list of items).

Table 2
*Example Motivated Strategies for Learning Questionnaire and Big Five Personality Test Questions*

| Motivated Strategies for Learning Questionnaire (7 likert scale) |
| --- |
| I prefer class work that is challenging so I can learn new things. |
| Compared to other students in this class, I expect to do well. |
| I am so nervous during a test that I cannot remember the facts that I have learned. |
| **Big Five Personality Test (5 pt likert scale)** |
| I see myself as someone who is talkative. |
| I see myself as someone who tends to find faults with others. |
| I see myself as someone who is reserved. |

Regarding the agent persona questionnaire, participants were asked to rate statements on a 6-point scale with 1 representing “strongly disagree”, 2 representing “somewhat disagree”, 3 representing “disagree”, 4 representing “agree”, 5 representing “somewhat agree” and 6 representing “strongly agree”. Participants rated various aspects of AutoTutor, including how engaging the system was (e.g., “I felt engaged during the tutoring session”), and how comparable the system was to an actual human tutor (“I felt that my interaction with AutoTutor was comparable to an interaction with a human tutor”). Appendix C presents the complete list of items.

Participants were asked to complete a short questionnaire that assessed various aspects of the confederate. There are possible extraneous variables when using confederates in any study, so the questionnaire was used in order to determine whether issues such as attractiveness and likeability might influence the results. Participants were asked to rate the confederate on a 6-point Likert scale on the following items: “I found the confederate to be believable", "I found the confederate to be attractive", "I found the confederate to be likeable", "I felt that the confederate was agreeable", "I felt that the confederate was able to answer my questions during the study", and "I felt that the confederate was knowledgeable." A complete list of items is presented in Appendix D.

*Procedure*. Students were trained on one of two versions of AutoTutor that were designed to test the impact of cognitive disequilibrium on questions and emotions. The study utilized a between-subjects design with random assignment to participants of the cognitive disequilibrium versus regular version of AutoTutor. Additionally, assignment of the computer literacy topics to tutor conditions was counterbalanced across all participants. For example, participant 1 might receive hardware, operating system, and internet in the *cognitive disequilibrium AutoTutor* condition. Participant 2 might receive operating system, internet, and hardware in the *regular AutoTutor* condition. The experiment consisted of three different phases: a pretest phase, a training phase, and a posttest phase.

*Pretest Phase*. Participants were tested individually during a two-hour session. First, participants completed three tests of individual differences: A demographic questionnaire, an 18-item domain knowledge questionnaire, and the Big Five Personality Test. Versions A vs. B of the domain knowledge questionnaire was counterbalanced with respect to the pre and posttest (see Appendix A for complete list of items). Second, the experimenter informed the participants that they would be learning about computer literacy from a computer-based tutor and would be tested after the tutoring session. Participants were told that the purpose of the study was to investigate learning using an intelligent tutoring system.

*Training Phase*. Participants were instructed to take a seat at the computer console where they were introduced to the confederate. The experimenter then instructed both the confederate and the participant on their roles for the experiment. The participants were always assigned the role of primary communicator. They were told that it was their job to collaborate with the confederate and to enter all input using the keyboard into AutoTutor. The confederate was always assigned to the role of secondary communicator. The confederates were told that it was their job to collaborate with the participant and assist in any way necessary (e.g., answering questions and technical problems). For the purpose of this study, the participants were lead to believe that the confederate was of equal status (i.e., a fellow undergraduate student completing the study for course credit).

Instructions for the session were read to the participants, including the general features of AutoTutor. All participants received the following instructions: "During this study you are going to interact with an intelligent tutoring system called AutoTutor. AutoTutor will help guide you through topics on computer literacy. During this session AutoTutor will ask you questions about computer literacy and your job will be to answer the questions to the best of your ability using the keyboard. The study will last approximately two hours. Please feel free to ask the experimenter any questions you may have at this time. The session will begin in a moment. Thank you for your participation!"

On the basis of random assignment, learners interacted with either (a) a version of AutoTutor that provided false feedback and false information (*cognitive disequilibrium AutoTutor)* or (b) a regular control version of AutoTutor (*regular AutoTutor).* Students' questions and facial action units were recorded throughout the duration of the learning session using a commercial webcam and standard desktop microphone. The students' facial action units were captured using a standard webcam. The webcam was positioned in a way to capture each participant's face at all times. Additionally, all dialogs between the participant and the confederate (including student questions) were recorded on the computer using a standard desktop microphone. The microphone was attached to the laptop which was directly in front of the participant at all times. The position of the microphone assured perfect recording quality.

Following the completion of the study, the experimenter reviewed all video and audio recordings to code for all student questions and any facial action units that arose. A second trained judge also coded questions and facial action units in order to assess inter-rater reliability. The second judge was considered trained regarding the facial action coding system once they scored an 80% or above on a test that accompanies the facial action coding system manual. The second judge was considered trained on question depth once they obtained a passing score (as judged by the experimenter) on an oral examination designed by the experimenter. Complete details regarding action unit and question coding are described in subsequent sections.

*Posttest Phase.* During the posttest phase the participants' knowledge of computer literacy was assessed using the corresponding version of either test A or B. Additionally, ratings of the confederate were collected in order to perform some assessment of aspects of the confederate that might influence participant engagement. For example, a male participant may ask more questions to a female confederate that he finds attractive. Alternatively, a participant may be more likely to ask a confederate questions if the participant perceives the confederate to be highly agreeable. To take into account these various potential confounds, participants rated the confederate on a number of different measures, such as believability and attractiveness (see Appendix D for a complete list of items).

The participants' question asking may also be influenced in part by how they perceive AutoTutor. Therefore, the participants completed an agent persona questionnaire (see Appendix C for a complete list of items) in which they were asked questions regarding their experience with the tutoring system. Because of the possibility that the confederate and the system might have an effect on question asking, motivation and learning strategies was also taken into account by having the participants complete the Motivated Strategies Learning Questionnaire. These tests in the posttest phases are elaborated below.

The posttest phase included four sets of questions. One set addressed domain knowledge, one addressed levels of motivation and attitudes towards AutoTutor, one addressed various aspects of the confederate, and one addressed various motivational aspects and learning strategies. Table 3 presents examples of the domain knowledge questions that tested participants' computer literacy knowledge through a different set of 18 multiple choice questions (see Appendix A for a complete list of items). The posttest was counterbalanced with the set of 18 multiple choice questions that was administered during the pretest phase. The post-interaction questionnaire asked participants to evaluate their tutorial session on measures of perceived performance, user satisfaction, motivation, feedback and task difficulty (9 questions). See Appendix C for a complete list of items. After the posttest was complete, participants were asked to complete the Motivated Strategies Learning Questionnaire (MSLQ) which consisted of a total of 81 questions (31 questions measuring motivation and 50 questions measuring learning strategies). See

Appendix B for a complete list of items. Following the MSLQ participants were asked to rate the confederate by completing the confederate perception questionnaire. See Appendix D for a complete list of items.

Participants were then debriefed and given the appropriate amount of course credit (approximately 2 hours). During debriefing, all participants that were randomly assigned to the *cognitive disequilibrium AutoTutor* condition were told that they received some incorrect information during their interaction with AutoTutor. To insure that participants in the *cognitive disequilibrium AutoTutor* condition did not leave the experiment with any misconceptions, they also received packets that contained a list of the false information that they received, accompanied with the correct content.

Table 3

*Example Pretest and Posttest Questions, Agent Persona Questions and Confederate Perception Questions*

| Computer Literacy Questions |
|---|
| If you install a sound card, why does your computer perform better? |
| (a) Because it can bypass the operating system when sound is needed |
| (b) Because it does not need the CPU |
| (c) Because sound no longer requires RAM |
| (d) Because there will be fewer bottlenecks when multitasking |
| What do the U.S. Postal System and the Internet have in common? |
| (a) They both use packet switching technology |
| (b) They are both controlled by the government in the USA |
| (c) The both use Uniform Resource Locators |
| (d) They both incorporate FIT and SLIP protocols |
| **Example Agent Persona Questionnaire (6 pt likert scale)** |
| I enjoyed interacting with AutoTutor. |
| I felt that my interaction with AutoTutor was comparable to an interaction with a human tutor. |
| I felt engaged during my interaction with AutoTutor. |
| I felt that AutoTutor was difficult to use and work with. |
| I felt that I learned new information from AutoTutor. |
| Understanding the material was important to me. |

(table continues)

Table 3 (continued)

| | |
|---|---|
| | While I was covering the material I tried to make everything fit together |
| | I felt that AutoTutor was knowledgeable. |
| | The feedback from AutoTutor was appropriate with respect to my progress. |
| Example Confederate Perception Questionnaire (6 pt likert scale) | |
| | I found the confederate to be believable. |
| | I found the confederate to be likeable. |
| | I felt that the confederate was knowledgeable. |

Each participant was randomly assigned to one of two conditions (*cognitive disequilibrium AutoTutor* versus *regular AutoTutor*). The order in which the participants received the subtopics (hardware, operating system, and internet) was counterbalanced across participants. Each participant interacted with two of the three possible subtopics. The third subtopic did not receive tutoring but served as a control to assess how much learning occurred when the student did not receive training on a particular topic. If the effectiveness of training is topic-specific, then there should be no learning evident for topics in which students received no tutoring. However, there might be some transfer from one topic to another; if so, then performance for this topic that received no tutoring should not differ from chance (i.e., 0 learning). Each participant interacted with their corresponding subtopics for a maximum of 30 minutes each. Students in all conditions entered their contributions via a keyboard.

Four confederates were used throughout the duration of the study. The confederates were undergraduate students enrolled for course credit at the University of Memphis. Before interacting with the participants, confederates were trained using a script developed by the tutoring research group at the University of Memphis. Confederates were told that all contributions must be entered by the participant via the keyboard. Confederates were instructed to answer any questions asked by the participants to the best of their ability. Furthermore, confederates sat across from the participants at all times but out of view of the computer screen, as illustrated in Figure 2.

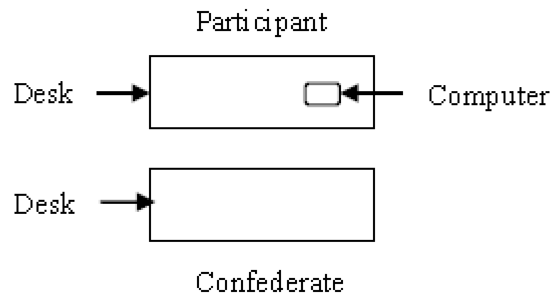


*Figure 2*. Position of the participant and confederate during testing phase.

*FACS and Question Coding*

All interactions during the tutoring sessions were recorded using a standard webcam and desktop microphone. Several properties of the session were observed and coded. First, two trained judges observed a sample of questions from the videos and determined when a question had been addressed to the confederate. The two trained judges collaborated in order to determine what exactly constituted a question. For example, any questions that were a repetition of the question just articulated by AutoTutor was not included in the analysis. Furthermore, the participant was required to make eye contact with the confederate in order for a question to be included. If the participant did not make eye contact with the confederate, the statement was coded as a "think aloud". Finally, any metacommunicative statements such as "What!?", "What did she say?" and "Did you understand her?" were not included as true information seeking questions. The judge rated the questions as either deep, intermediate, or shallow based on the Graesser and Person (1994) question taxonomy, which is presented in Appendix E.

Second, the judge observed the videos and coded in parallel each individual's action units. The judgments referred to time spans either 3 seconds before a question was asked, during the question, versus 3 seconds after a question had been asked. More specifically, the focus was on action units that have been shown to be associated with both confusion (4 and 7) and frustration (1, 2, and 14). A second judge rated all facial action units and student questions in order to obtain inter-rater reliability. Judges were trained to detect facial action units using the Facial Action Coding System (Ekman & Friesen, 1978). As previously mentioned, judges were considered sufficiently trained after they had obtained a score of 80% or higher on the practice test that accompanies the facial action coding system manual. Judges were trained in question depth (i.e., deep, intermediate, and shallow) using the Graesser and Person (1994) question taxonomy.

# RESULTS

The 48 participants in the study completed two AutoTutor subtopics within their assigned condition and completed all pretest and posttest questions. In this Results section we explored the differences between the two conditions on: (1) quantity of questions, (2) quality of questions, (3) learning gains, (4) presence of the affective/cognitive state of confusion preceding, during, and following a question, and (5) individual difference measures as predictors of student generated questions.

## *Quantity of Questions*

All student-generated questions were transcribed and recorded in order to explore any significant differences between participants in the two conditions. A total of 2,107 questions were generated by the participants. Results revealed a significant difference in the amount of questions asked between the participants in the *regular AutoTutor* condition and the participants in the *cognitive disequilibrium AutoTutor* condition, $F(1,45) = 4.22$, $p < .05$, eta-squared = .087. The means and standard deviations for the raw amount of questions asked per student during the learning session are shown in Table 4. It should be noted that this result is opposite to the prediction of the existing cognitive disequilibrium literature, which states that inducing cognitive disequilibrium should increase the quantity of student generated questions.

Table 4

*Means and Standard Deviations of Raw Number of Questions in the Two AutoTutor Conditions*

| | Regular AutoTutor (*n*=24) | Cognitive Disequilibrium AutoTutor (*n*=24) |
|---|---|---|
| Quantity of Questions | 52.62 (30.98) | 35.16 (26.55) |

One potential reason that students did not ask as many questions in the *cognitive disequilibrium AutoTutor* condition is that they may have lost confidence in AutoTutor as a reliable information source because it had flawed information. If so, then questions would be expected to decrease over time in that condition, but remain comparatively steady in the *regular AutoTutor* condition. This would yield a condition by time interaction in the quantity of questions. To test this, a 2 (condition) X 4 (time intervals) mixed ANOVA was conducted on potential differences in *when* students were asking questions during their learning session. Each participant's interaction with AutoTutor was divided into four time intervals. Because some participants did not interact with AutoTutor for the total allotted time (60 minutes) the vincentization method was used to calculate time intervals. For example, the total session time for each participant was calculated individually and divided by four in order to determine the length of each time interval. Results showed no significant differences in questions asked across all four time periods. However, it warrants mentioning that the largest gap in question asking occurred during the fourth time interval, although it was not significant. Figure 3 shows the mean number of questions asked for the two AutoTutor conditions as a function of the four time intervals.

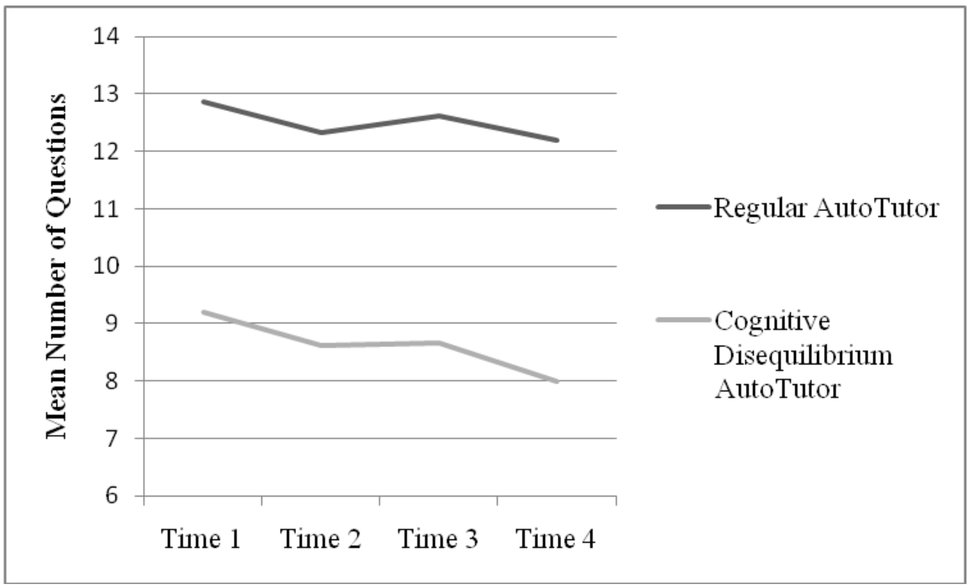


*Figure 3.* Number of Questions Asked for Two AutoTutor Conditions as a Function of Four Time Intervals

*Quality of Questions*

In order to examine potential differences in the quality of questions that participants asked between the two conditions, two independent judges rated each transcribed question and coded it as shallow, intermediate, vs. deep based on the Graesser and Person (1994) taxonomy. An example of a shallow question from the corpus was "Is it the CPU?". An example intermediate question from the corpus was "What does file transfer protocol mean?" and an example deep question from the corpus was "How exactly are RAM and the hard disk the computer's hard drive?". A 2 (condition) X 3 (deep, intermediate, and shallow) mixed ANOVA was conducted in order to explore any possible differences in question quality between the two conditions. Results revealed no significant differences in the number of deep questions asked between participants in the two conditions, $F$ (1, 46) = .071, $p$ = .79, eta-squared = .002. Furthermore, no significant differences were found between participants in the two conditions in the amount of intermediate questions asked during the learning session, $F$ (1, 46) = 1.409, $p$ = .241, eta-squared = .030. However, a significant difference was found in the amount of shallow questions that were asked by the participants in the two conditions, $F$ (1, 46) = 3.90, $p$ = .054, eta-squared = .078. The means and standard deviations of the three question categories can be seen in Table 5.

Table 5

*Means and Standard Deviation for Question Depth in the Two AutoTutor Conditions*

| Measures | Regular AutoTutor (*n*=24) | Cognitive Disequilibrium AutoTutor (*n*=24) |
|---|---|---|
| Deep Questions | .250 (.442) | .291 (.624) |
| Intermediate Questions | 2.04 (2.86) | 1.25 (1.56) |
| Shallow Questions | 47.87 (27.36) | 32.95 (24.88) |

*Learning Measures*

Proportional learning gains were computed to assess knowledge gains during the learning session. To obtain the proportion correct for each student (on pretest and posttest), the number of correct items on the pretest and posttest were divided by the total number of questions. Participants in both conditions did improve from pretest to posttest, but there were no significant differences in learning between the participants in the two conditions, $F$ (1, 46) = .382, $p$ = .54, eta squared = .008. Additionally, results revealed that there were no significant differences between conditions on the pretest. Means and standard deviations are presented in Table 6.

Table 6

*Means and Standard Deviation for Learning Gains in the Two AutoTutor Conditions*

| | Regular AutoTutor (*n*=24) | Cognitive Disequilibrium AutoTutor (*n*=24) |
|---|---|---|
| Proportion Gain Scores | .175 (.211) | .135 (.232) |

*Action Units*

The purpose of this analysis was to investigate whether a temporal relationship existed between student generated questions and when they experienced cognitive disequilibrium. All action units pertaining to confusion and frustration were coded three seconds prior to a student generated question, during a student generated question, and three seconds following a student generated question. A 2 (condition) X 3 (before, during, or after) mixed ANOVA was computed on the incidence of action units within a time interval. The occurrences of confusion (action units 4 and 7) and frustration (action units 1, 2, and 14) can be seen in Table 7.

*Table 7*

Means (Standard Deviation) on Action Units Surrounding Student Generated Questions

| Action Units | Regular AutoTutor | Cognitive Disequilibrium AutoTutor |
|---|---|---|
| AU1, 2, & 14 Before (Frustration) | .016 (.067) | .001 (.004) |
| AU1, 2, & 14 During (Frustration) | .003 (.014) | .000 (.002) |
| AU1, 2, & 14 After (Frustration) | .016 (.068) | .001 (.004) |
| AU4 & 7 Before (Confusion) | .159(.214) | .150 (.211) |
| AU4 & 7 During (Confusion) | .190 (.192) | .156 (.170) |
| AU4 & 7 After (Confusion) | .120 (.173) | .132 (.148) |

No significant differences were found between AutoTutor conditions in the occurrence of any individual action unit or combination of action units preceding a question, during a question, or following a question. Table 7 presents means and standard deviation for action units surrounding student generated questions.

One set of analyses concentrated on action units that were expected to be predictive of the emotion frustration (i.e., action units 1, 2, and 14). Results revealed no significant difference between AutoTutor conditions in the occurrence of action unit 1 (Inner Brow Raiser) before a question, $F$ (1, 46) = .070, $p$ = .79, eta squared = .002, during a question $F$ (1, 46) = .056, $p$ = .81, eta squared = .001, or following a question $F$ (1, 46) = .254, $p$ = .61, eta squared = .005. Additionally, results revealed a significant main effect as a function of time, $F$ (1, 46) = 13.043, $p$ < .01. More specifically, there was a significantly greater proportion of action unit 1 occurring during a question ($M$ = .28) rather than before a question ($M$ = .13) , $p$ < .001. There was a significantly greater amount of action unit 1 occurring during a question ($M$ = .28) rather than after a question ($M$ = .20),  $p$ < .001. Lastly, there was a significantly greater proportion of action unit 1 occurring after a question (M = .20) rather than before a question ($M$ = .13),  $p$ < .01. No significant condition by time interaction was present.

A second set of analyses examined any potential differences between the two conditions regarding action unit 2 (Outer Brow Raiser). Results revealed no significant differences in the occurrence of action unit 2 before a question $F$ (1, 46) = .449, $p$ = .50, eta squared = .010, during a question $F$ (1, 46) = .048, $p$ = .82, eta squared = .001, or following a question $F$ (1, 46) = .258, $p$ = .61, eta squared = .006. Additionally, results revealed a significant main effect as a function of time, $F$ (1, 46) = 18.101, $p$ < .001.

More specifically, there was a greater proportion of action unit 2 occurring during a question (M = .29) rather than before a question (*M* = .14) a question, $p < .001$. A significantly greater proportion of action unit 2 was found to occur during a question (*M* = .29) rather than after a question (*M* = .19), $p < .001$. Lastly, there was a significantly greater proportion of action unit 2 occurring after a question (*M* = .19) rather than before a question (*M* = .14), $p < .001$. No significant condition by time interaction was discovered.

Results exploring the differences between the two conditions regarding action unit 14 (Dimpler) revealed no significant differences before a question $F$ (1, 46) = 1.160, $p$ = .28, eta squared = .025, during a question $F$ (1, 46) = .748, $p$ = .39, eta squared = .016, or following a question $F$ (1, 46) = 2.125, $p$ = .15, eta squared = .044. Results revealed no significant main effect as a function of time for action unit 14, $F$ (1, 46) = .936, $p$ = .34. Additionally, no significant condition by time interaction was found.

Furthermore, we examined any potential differences using the combination of action units 1 (Inner Brow Raiser), 2 (Outer Brow Raiser), and 14 (Dimpler). Results revealed no significant differences between the combination of action units 1, 2, or 14 before a question $F$ (1, 46) = 1.143, $p$ = .29, eta squared = .024, during a question $F$ (1, 46) = 1.136, $p$ = .29, eta squared = .024, or following a question $F$ (1, 46) = 1.168, $p$ = .28, eta squared = .025. Additionally, results revealed no significant main effect as a function of time for action units 1, 2, and 14, $F$ (1, 46) = .051, $p$ = .82. No significant condition by time interaction was found.

Analyses were conducted exploring potential differences between the two conditions regarding the action units that accompany the emotion/affective state of confusion (i.e., 4 and 7). First, we examined any potential differences between the two conditions regarding action unit 4 (Brow Lowerer). Results revealed no significant difference in the occurrence of action unit 4 before a question, $F$ (1, 46) = .207, $p$ = .65, eta squared = .004, during a question, $F$ (1, 46) = 1.075, $p$ = .30, eta squared = .023, or following a question, $F$ (1, 46) = .001, $p$ = .97, eta squared = .000. Additionally, results revealed a significant main effect as a function of time, $F$ (1, 46) = 3.532, $p$ < .05. More specifically, a significantly greater proportion of action unit 4 occurred during a question ($M$ = .21) rather than after a question ($M$ = .16), $p$ < .01. No significant condition by time interaction was found.

Secondly, we examined the differences between the two conditions pertaining to the occurrence of action unit 7 (Lid Tightner). Results showed no significant differences between the two conditions on action unit 7 before a question, $F$ (1, 46) = .068, $p$ = .79, eta squared = .001, during a question, $F$ (1. 46) = .346, $p$ = .55, eta squared = .007, or following a question, $F$ (1, 46) = .172, $p$ = .68, eta squared = .004. Results revealed no significant main effect as a function of time for action unit 7, $F$ (1, 46) = 2.141, $p$ = .12. No significant condition by interaction was found.

Finally, we examined any differences between the two conditions regarding the combination of action units that accompany the emotion/affective state of confusion (i.e., 4 and 7). Results revealed no significant differences before a question, $F$ (1, 46) = .019, $p$ = .89, eta squared = .000, during a question, $F$ (1, 46) = .346, $p$ = .55, eta squared = .007, or following a question, $F$ (1, 46) = .066, $p$ = .79, eta squared = .001. Results revealed a significant main effect as a function of time, $F$ (1, 46) = 3.646, $p$ < .05. More specifically, a significantly greater proportion of action units 4 and 7 were discovered to occur during a question (*M* = .18) rather than after a question (*M* = .13), $p < .01$. No significant condition by interaction was found.

A paired-samples *t*-test revealed a significantly greater amount of confusion than frustration before a question $t$ (47) = 4.814, $p <.001$, during a question $t$ (47) = 6.623, $p < .001$, and after a question $t$ (47) = 4.774, $p < .001$. Frustration was indeed rarely manifested in the facial expressions. This result is consistent with earlier research on the relationship between frustration and the action units (D'Mello & Graesser, 2010).

*Individual Differences Measures*

Because one focus of this study was related to confusion and frustration, multiple regression analyses were conducted in order to determine whether the constructs on the agent persona questionnaire that related to confusion and frustration were significant predictors of quantity of questions asked. Results revealed no significant results for either frustration $t$ (45) = .817, $p = .41$, $R^2 = .05$ or confusion $t$ (45) = .511, $p = .61$, $R^2 = .05$. No other items on the agent persona questionnaire were significant predictors of questions generated. Further analyses revealed a significant difference between the two conditions on the agent persona questionnaire item "While I was covering the material, I tried to make everything fit together", $t$ (46) = 2.197, $p < .05$. More specifically, participants in the *regular AutoTutor* condition rated significantly higher than did the participants in the *cognitive disequilibrium* condition. Additionally, the two conditions significantly differed on the item "AutoTutor's emotions were natural" with the participants in the *regular AutoTutor* condition rating higher, $t$ (46) = 2.846, $p < .01$. Lastly, there was a significant difference between the two conditions on the item "The feedback from AutoTutor was appropriate with my respect to my progress" with participants in the *regular AutoTutor* conditions rating the system higher, $t$ (46) = 2.751, $p < .01$.

There was a possibility that the confederate may have had an influence on the questions generated by the students. To assess this, a regression analysis was conducted on the overall perception of the confederate and the number of questions asked during the learning session. In order to obtain an overall perception score of the confederates, each item on the confederate perception questionnaire was totaled and divided by the total number of items available. The regression analysis revealed that overall confederate perception was a significant predictor of questions asked, $t$ (46) = 2.238, $p < .05$, $R^2$ = .09. A one-way ANOVA was conducted in order to determine if there were any significant differences between the two conditions and the overall perception of the confederates. Results revealed a marginally significant difference suggesting that participants in the *regular AutoTutor* condition ($M$ = 5.15) had a better perception of the confederate than the participants in the *cognitive disequilibrium AutoTutor* condition, ($M$ = 4.78), $F$ (1,46) = 3.614, $p$ = .06, eta-squared = .073.

All participants completed the Motivated Strategies Learning Questionnaire (MSLQ). A regression analysis was conducted on the main two constructs measured by the MSLQ (motivation and learning strategies). In the current multiple regression analysis the predictor variables included motivation and learning strategies with number of questions as the criterion variable. The items related to the motivation construct included: Intrinsic Goal Orientation, Extrinsic Goal Orientation, Task Value, Control of Learning Beliefs, Self Efficacy for Learning and Performance, and Test Anxiety. These items were totaled giving us an overall motivation score for each participant. The results of the regression analysis revealed that motivation was a significant predictor of questions asked, *t* (46) = 3.079, $p < .01$, $R^2 = .17$. Additionally, results revealed no significant differences between the two conditions regarding motivation, *t* (46) = .379, $p = .70$. The items related to the learning strategies construct included: Rehearsal, Elaboration, Organization, Critical Thinking, Metacognitive Self-Regulation, Time and Study Environment, Effort Regulation, Peer Learning, and Help Seeking. These items were totaled giving us an overall learning strategies score for each participant. Results showed that learning strategies were not a significant predictor of question asking *t* (46) = 1.543, $p = .13$, $R^2 = .04$. Additional results revealed no significant differences between the two condition on learning strategies, *t* (46) = 1.761, $p = .08$.

A multiple regression analysis was conducted with the personality traits "agreeableness" and "neuroticism" as predictor variables and number of questions asked as a criterion variable to test whether any personality traits were significant predictors of student generated questions. Results revealed that the personality traits "agreeableness", *t* (45) = 2.357, $p < .05$ and "neuroticism", *t* (45) = 2.531, $p < .05$ were significant predictors of questions asked, $R^2 = .17$. Further results revealed a significant difference between the two conditions regarding the personality trait "agreeableness", *t* (46) = 2.891, $p < .01$. Additionally, results revealed a significant difference between the two conditions regarding the personality trait "extraversion", *t* (46) = 2.044, $p < .05$.

# DISCUSSION

The results from this study  sheds some additional light on whether it is possible to increase the amount of questions that students ask while learning about computer literacy from an intelligent tutoring system. The results from this study discovered a significant difference in the quantity of questions asked between the two conditions. Additionally, no significant differences were found between the two conditions on deep or intermediate questions. Results also showed that there were no significant differences in the manifestation of certain action units know to accompany specific emotions (i.e., confusion and frustration) before a question, during a question, or after a question. Furthermore, we found no significant differences in learning gains between the two conditions. Lastly, results revealed that certain measures of individual differences are a significant predictor of number of questions generated.

As Graesser and Person (1994) pointed out, a student in a typical setting such as a classroom asks only .11 questions per hour. Furthermore, it has been shown that even in one-to-one tutoring sessions students still typically only ask approximately 22.7 questions per hour. Although this increase in student question asking is encouraging, there is no doubt room for improvement. The purpose of this study was to explore the possible effects of inducing cognitive disequilibrium in participants while interacting with AutoTutor.

The results showed that participants in both conditions showed an increase in question asking compared to typical classroom settings and one-to-one human tutoring. More specifically, the participants in the *regular AutoTutor* condition asked on average 53 questions per hour and the participants in the *cognitive disequilibrium AutoTutor* condition asked on average 35 question per hour. However, it was predicted that students who were placed in a state of cognitive disequilibrium would generate significantly more questions. The justification for this prediction was that it was thought that students who were in a state of cognitive disequilibrium would be forced to self question until equilibrium was restored.  The results from this study are contradictory to the previous results discovered in the cognitive disequilibrium literature. For example, Graesser and Olde (2003) discovered presenting students a breakdown of an everyday device was one effective way of inducing cognitive disequilibrium, which in turn influenced student question generation. This result was not confirmed in the present dissertation when AutoTutor disseminated false feedback and information to the student learner.

One potential explanation for these counter-intuitive findings can be attributed to the confederates. Each confederate was told to answer each student generated question to the best of their ability. However, the students that were randomly assigned to the cognitive disequilibrium version of AutoTutor received a small percentage (25%) of incorrect feedback during their session. Of course, the confederates never knew when AutoTutor would deliver the incorrect feedback and therefore attempted to answer participants' questions to the best of their ability. Unfortunately, due to AutoTutor's incorrect feedback to a correct confederate response, over time the participants may have "lost faith" in the confederate's ability to correctly answer their question. If the participant no longer felt that the confederate was able to answer the question, this is a violation of principles three, eight and ten of question asking according to Van Der Meij (1987). Van Der Meij posits that for a true information seeking question to occur, eleven conditions must be met. These 11 conditions are as follows:

1) The questioner does not know the information asked for with the question

2) The question specifies the information sought

3) The questioner believes that the presuppositions of the question are true

4) The questioner believes that an answer exists

5) The questioner wants to know the answer

6) The questioner can assess whether a reply constitutes an answer

7) The questioner poses the question only if the benefits exceed the costs

8) The questioner believes that the respondent knows the answer

9) The questioner believes that the respondent will not give the answer in absence of a question

10) The questioner believes that the respondent will supply the answer and

11) A question solicits a reply

The idea of participants "losing faith" in the confederate is also supported by the results of the regression analyses exploring overall confederate perception. For example, results from the regression analysis revealed that overall confederate perception was in fact a significant predictor of questions asked during the learning session. Furthermore, the results from the one-way ANOVA showed a marginally significant difference between the two conditions regarding the overall confederate perception. In other words, the results showed that participants in the *cognitive disequilibrium AutoTutor* condition had a lower overall perception of the confederate than did the participants in the *regular AutoTutor* condition. This difference could possibly explain the significant difference in the raw amount of questions asked between the two conditions.

Another potential explanation as to why participants in the *regular AutoTutor* condition asked significantly more questions during the learning session may be due to the content of the tutoring session. As mentioned earlier, one technique that was used in attempting to induce cognitive disequilibrium in the participants was by providing a small percentage of false information during their interaction with AutoTutor. However, the content that was provided to each participant during their tutoring session covered information regarding computer literacy. It may be that participants simply did not have the motivation or interest in computer literacy to notice when obvious incorrect information was presented (e.g., "Without RAM, it is impossible to run *marathons*"). This idea is supported by the results of the regression analysis that indicated that motivation was a significant predictor of question generation. Perhaps if we had used a more controversial topic such as global warming or the death penalty participants would be more likely to notice discrepant information when presented.

Additionally, the results of this study revealed no significant differences in the amount of deep or intermediate questions asked between the participants in the two conditions. This was an unexpected finding. For example, Graesser and Olde placed students in a state of cognitive disequilibrium by giving them a breakdown scenario of an everyday device. The results of their study showed that the most robust predictor of student understanding was not the raw number of questions asked but the quality of the questions asked. It was predicted in the current study that after participants in the *cognitive disequilibrium AutoTutor* condition received anomalous information it would cause them to think at a deeper level about the material which in turn would cause them to generate more deep and intermediate level questions. It is thought that the predicted finding would be discovered if the measure of the proportion of questions that were deep or intermediate were more prominent. However, there were so few deep and intermediate questions that such an analysis becomes difficult to interpret.

The dissertation also investigated the temporal relationship between cognitive disequilibrium/confusion and student generated questions. All action units related to confusion and frustration were coded three seconds prior to a question, during a question, and three seconds following a question. Results revealed that participants in both conditions showed a comparable amount of cognitive disequilibrium (as measured by facial expressions) before generating a question, during a question and following a question with no significant differences between the two conditions. Unfortunately, these results are not consistent with the research conducted by Graesser, Lu, Olde et al. (2005). Graesser and colleagues explored the effects of cognitive disequilibrium and student generated questions as a function of time. During their study, the researchers tracked the eye movements of participants while they read a text that described the breakdown of an everyday device. It was hypothesized that if cognitive disequilibrium does in fact trigger questions, then participants should have fixated (based on eye tracking) on the faulty parts of the device prior to or during a question. Results revealed that the good comprehenders did fixate significantly more on the faulty locations both *before* and *during* question generation. Our results showed that participants in both the *regular AutoTutor* condition and the *cognitive disequilibrium AutoTutor* condition showed no difference in the occurrence of cognitive disequilibrium (as measured by facial action units) and question generation.

Nevertheless, there is a plausible explanation as to why we found these null results. The attempt to induce cognitive disequilibrium (i.e., false feedback and false information) may have put students in a state of frustration rather than confusion. This motivated the coding of all action units that related to the emotion of frustration. However, the results revealed a very small proportion of frustration actually occurred throughout the study, at least those that could be detected by the facial action units.

Moreover, the results of a paired samples *t*-test showed significantly higher proportions of confusion surrounding each question rather than frustration. These results suggest that our methods of inducing cognitive disequilibrium were actually inducing cognitive disequilibrium more than frustration.  It should be noted that frustration is normally not detected in the face, but is instead detected from the dialogue or body posture according to available research that has investigated emotion sensing from multiple channels (D'Mello & Graesser, 2010).  From this perspective, it is not surprising that facial actions associated with frustration were so low in frequency.

It is possible that the present methods of inducing cognitive disequilibrium may not have been an effective manipulation in inducing cognitive disequilibrium in the participants. This possibility is somewhat supported in the relatively low amounts of cognitive disequilibrium recorded in both the *regular AutoTutor* condition (16%, 19%, and 12%) and the *cognitive disequilibrium AutoTutor* condition (15%, 16%, and 13%). Future studies need to continue to explore other possibilities of effectively inducing cognitive disequilibrium.

The social dimensions of the present study may be one reason that cognitive disequilibrium was not preceding or following questions. Perhaps the participants did experience a high amount of cognitive disequilibrium in either one or both of the conditions, but they were stifled due to social pressures (e.g., afraid of looking ignorant in front of a stranger). Future analyses need to take in consideration all occurrences of cognitive disequilibrium throughout the learning session (as manifested in facial action units 4 and 7) and determine what proportions of questions are actually being articulated when these instances of confusion occur.

Participants in this study were also given multiple-choice tests in order to determine which version of AutoTutor would produce the highest learning gains. Results showed that there were no significant differences in learning between the two conditions. If anything, participants in the *regular AutoTutor* condition leaned to having higher posttest learning scores. Although all participants in the *cognitive disequilibrium AutoTutor* condition received all corrected information in the summary before moving to the next topic, this contradictory information may have been confusing to the participants and thereby lowered learning gain scores.

Participants in both conditions gained a percentage of learning from pretest to post test (18% and 14%). These gain scores may in fact be one possible explanation for the low number of deep and intermediate questions asked by participants in both conditions. As mentioned earlier, Graesser and Olde (2003) discovered that the “best” questions came from the deep comprehenders. Although, participants are showing an improvement from pretest to post test they may just be comprehending the material which they covered in the pretest and not actually acquiring deep knowledge, which in turn could explain the low number of deep and intermediate questions.

This study does have some limitations that warrant mentioning. The first limitation involves the lack of control of the confederate during the participants’ interaction with AutoTutor. For example, each confederate was told to answer any participant question to the best of their ability. However, during the session some participants initiated a conversation with the confederates. Because confederates were not specifically told how to handle casual conversation with the participants, this may have had an influence on the quality and quantity of questions. Future research needs to develop a more rigorous script for the confederates to follow during each individual learning session.

A second limitation involves the transcription of the questions. In order to determine what constituted a genuine question, the experimenter and a trained judge collaborated using a sample of questions from the corpus in order to develop rules that specify which questions should be included. Contributions from the participants that we did not consider genuine questions (e.g., “What!?”) may be included by other researchers with different interpretations. Future research that explores student question generation needs to use existing guidelines, such as the Van der Meij (1998) principles of question asking, in order to more accurately decide what is considered a good, genuine information seeking question.

*References*

Baker, L. (1979). Comprehension monitoring: Identifying and coping with text confusions. *Journal of Reading Behavior, 11*, 363-374.

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

Bloom, B.S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive Domain*. New York: McKay.

Breazeal, C. (2003). *Designing sociable robots* .Cambridge: MIT Press.

Collins, A., Brown, J. S., & Larkin, K. M. (1980). Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.) *Theoretical issues in reading comprehension* (pp. 385-407). Hillsdale, NJ: Erlbaum.

Conati C. (2002). Probabilistic assessment of user's emotions in educational games. *Journal of Applied Artificial Intelligence, 16*, 555-575.

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote-Aloud during Learning with AutoTutor: Applying the Facial Action Coding System to Affective States during Learning. *Cognition and Emotion, 22*, 777 – 788

Craig, S.D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media, 29*, 241-250.

Craig, S. D., VanLehn, K. & Chi, M. T. H. (in press). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of Educational Psychology*.

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

Davey, B., & McBride, S. (1986). Effects of question generating training on reading comprehension. *Journal of Educational Psychology, 78*, 256-262.

De Vicente, A., & Pain, H. (2002). Informing the detection of students' motivational state: An empirical study. In S.A. Cerri, G. Gouarderes, & F. Paraguacu (Eds.), *Intelligent tutoring systems 2002*. Berlin, Germany: Springer.

Dillon, J. T. (1988). *Questioning and teaching: A manual of practice*. New York: Teachers College Press.

Ekman, P., & Friesen, W. V. (1978). *The facial action coding system: A technique for the measurement of facial movement*. Palo Alto: Consulting Psychologists Press.

Fishbein, H. D., Eckart, T., Lauver, E., Van Leeuwen, R., & Langmeyer, D. (1990). Learners' questions and comprehension in a tutoring setting. *Journal of Educational Psychology, 82*, 163–170.

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology*, *10*, 17-32.

Graesser, A.C., Chipman, P., Haynes, B.C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions in Education,* 48, 612-618.

Graesser, A. C., D'Mello, S. K., Chipman, P., King, B., & McDaniel, B. (2007). Exploring relationships between affect and learning with AutoTutor. In R. Luckin, K. Koedinger, & J. Greer (Eds.), *Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work* (pp. 16–23). Amsterdam: IOS Press.

Graesser, A.C., D'Mello, S.K., Craig, S.D., Witherspoon, A., Sullins, J., McDaniel, B., & Gholson, B. (2008). The relationship between affect states and dialogue patterns during interactions with AutoTutor. *Journal of Interactive Learning Research, 19*, 293-312.

Graesser, A.C., Jackson, G.T., & McDaniel, B. (2007). AutoTutor holds conversations with learners that are responsive to their cognitive and emotional states. *Educational Technology, 47*, 19-22.

Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., & Louwerse, M.M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 180-193

Graesser, A.C., Lu, S., Olde, B.A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory and Cognition*, *33*, 1235-1247.

Graesser, A. C., & McMahen, C. L. (1993). Anomalous information triggers questions when adults solve problems and comprehend stories. *Journal of Educational Psychology, 85*, 136-151.

Graesser, A.C., & Olde, B.A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology, 95,* 524-536.

Graesser, A. C., Ozuru, Y., & Sullins, J. (2009). What is a good question? In M. G. McKeown & L. Kucan (Eds.), *Threads of coherence in research on the development of reading ability* (pp. 112-141). NY: Guilford.

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

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 3,* 371-395.

Graesser, A. C., Witherspoon, A., McDaniel, B., D'Mello, S., Chipman, P., & Gholson, B. (2006). Detection of emotions during learning with AutoTutor. In R. Son (Ed.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 285–290). Mahwah, NJ: Erlbaum.

Heilman, M., & Smith, N. A. (2010). Extracting Simplified Statements for Factual Question Generation. In K. E. Boyer & P. Piwek (Eds.), *QG2010: The Third Workshop on Question Generation* (11-20). *Springer:Berlin / Heidelberg*.

Jackson, G.T., Mathews, E.C., Lin, D., & Graesser, A.C. (2003). Modeling Student Performance to Enhance the Pedagogy of AutoTutor. In P. Brusilovsky, A. Corbett, & F. de Rosis (Eds.), *Lecture Notes in Artificial Intelligence: 2702* (pp. 368-372). New York: Springer.

Kort, B., Reilly, R., & Picard, R. (2001). An affective model of interplay between emotions and learning: Reengineering educational pedagogy—building a learning companion. In T. Okamoto, R. Hartley, Kinshuk, & J. P. Klus (Eds.), *Proceedings IEEE International Conference on Advanced Learning Technology: Issues, Achievements and Challenges* (pp. 43-48). Madison, WI: IEEE Computer Society.

Lepper, M. R., & Woolverton, M. (2002). The wisdom of practice: Lessons learned from the study of highly effective tutors. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 135-158). Orlando, FL: Academic Press.

Litman, D. J., & Forbes-Riley, K. (2004). Predicting student emotions in computer-human tutoring dialogues. In *Proceedings of the 42nd annual meeting of the association for computational linguistics (pp. 352-359).* East Stroudsburg, PA: Association for Computational Linguistics.

Macaulay, D. (1988). *The way things work.* Boston: Houghton Mifflin.

McDaniel, B. T., D'Mello, S. K., King, B. G., Chipman, P., Tapp, K., & Graesser, A. C. (2007). Facial Features for Affective  State Detection in Learning Environments. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 467-472). Austin, TX: Cognitive Science Society.

Miyake, N., & Norman, D.A. (1979). To ask a question, one must know enough to know what is not known. *Journal of Verbal Learning and Verbal Behavior, 18*, 357-364.

Otero, J., & Graesser, A.C. (2001). PREG: Elements of a model of question asking. *Cognition & Instruction, 19,* 143-175.

Picard, R. W. (1997). *Affective computing*. Cambridge: MIT Press.

Pintrich, P. R., Smith, D., Garcia, T., and McKeachie, W. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*, The University of Michigan, Ann Arbor, MI.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*, 181-221.

Sternberg, R. J. (1987). Questioning and intelligence. *Question Exchange*, *1*, 11-13.

Sullins, J., & McNamara, D.S. (2009). iSTART question training module: Training students efficient questioning skills. Presented at the 2nd Workshop on Question Generation, Brighton, United Kingdom.

Van der Meij, H. (1988). Constraints on question asking in classrooms. *Journal of Educational Psychology, 80*, 401-405.

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.

Wang, N., Johnson, W.L., Mayer, R., Rizzzo, P., Shaw, E., & Collins, H. (2005). The politeness effect: Pedagogical agents and learning gains. In C. Looi, G. McCalla, B. Bredeweg, & J. Breuker, (Eds.), *Artificial intelligence in education* (pp. 686—693). Amsterdam: IOS Press.

Wisher, R. A., & Graesser, A. C. (2007). Question asking in advanced distributed learning environments. In S. M. Fiore & E. Salas (Eds.), *Toward a science of distributed learning and training* (pp. 209–234). Washington, DC: American Psychological Association.

**Appendix A: Computer Literacy Test Questions**

**Pre/Posttest Form A**

**General instructions:** Thank you for participating in our study. Please answer all questions as best you can. If you have any questions while completing this assessment, please ask the experimenter.

Why is RAM considered to be volatile memory?

a. **Its contents can be changed**.
b. Its contents are easily lost.
c. The CPU can read its contents.
d. It can send error messages to computer users.

What does the CPU use RAM for when executing an application program?

a. Long term storage.
b. To bypass the operating system.
c. **Short term storage**.
d. To hold the instructions necessary to reboot.

What will happen if the CPU failes to perform mathematics?

a. It will not communicate with peripherals.
b. It will only process some of the data.
c. **It will not process any of the data**.
d. It will crash the operating system.

How does the CPU obtain information from the keyboard?

a. Information goes through RAM to the CPU.
b. Information goes from the monitor to RAM to the CPU.
c. Information goes directly from the monitor to the CPU.
d. **Information goes directly from the keyboard to the CPU.**

Why does a computer process data stored on a cassette more slowly than data stored on a floppy disk?

a. Because cassette tapes are longer.
b. **Because cassettes are not random access devices**.
c. Because floppies are smaller in size.
d. Because floppies cannot hold as much data.

What do cassettes and compact disks have in common?

a. They usually lose data after a year.
b. They are both random access devices.
c. They cannot have information overwritten once it is recorded.
**d. They can both store large amounts of information.**

Why is the Internet said to be a new society or new world?

a. **Because it has its own rules of etiquette**.
b. Because it has penetrated practically every country in the world.
c. Because people interact differently in electronic media.
d. Because it is such a rapid means of communication.

How would you know if an e-mail message was a violation of basic rules that apply to the internet?

a. It used symbols to convey emotions.
b. It was not from someone you know.
c. **It was not short and to the point**.
d. It was not an advertisement.

What happens if Telnet is not available?

a. **You cannot do remote login**.
b. You cannot transfer files.
c. You cannot transfer files as attachments.
d. You cannot use e-mail at all.

What would happen if FTP was not available?

a. You cannot do remote login.
b. **You cannot transfer files**.
c. You cannot transfer any protocols.
d. You cannot use e-mail at all.

How do you communicate over the internet if you use terminal emulation?

a. Through a LAN
b. **Standard phone lines**.
c. Using SLIP software.
d. Use an FTP connection.

How would you connect to the internet if you wanted a temporary but full connection?

a. Through a LAN.
b. **SLIP software**.
c. An on-line service host.
d. Use an FTP connection.

If your computer system is failing to do certain maintenance tasks, what is probably responsible?

a. The hard drive is faulty.
b. The CPU has malfunctioned.
c. **A utility program has malfunctioned**.
d. The operating system has malfunctioned.

When files are damaged due to ordinary use, what does the computer system use to repair them?

a. Virus protection software.
b. BIOS.
c. **Utility programs**.
d. The operating system.

When you first turn on a computer what is the most important consequence?

a. It contacts the keyboard and monitor.
b. It sends a message to RAM and ROM.
c. **It loads the operating system**.
d. It sends a message to the hard drive and the operating system.

What is the disadvantage of having the operating system stored in ROM?

a. ROM is slow relative to RAM.
b. The CPU has more difficulty locating ROM than RAM.
c. **It is hard to upgrade**.
d. The hard drive is easier for the CPU to access than ROM.

When a person is using an application, where is it in relation to the RAM, the ROM, the operating system, the hard drive, the interface, and the person?

a. It is between the person and the interface.
b. **It is between the interface and the operating system**.
c. It is between the person and RAM.
d. It is between the operating system and RAM.

After a computer is booted up, where is the operating system in relation to the interface, a running application, and the computer hardware?

a. It is between the person and the interface.
b. It is between the interface and the running application.
c. **It is between the application and the computer hardware**.
d. It is still located in ROM.

## Pre/Posttest Form B

**General instructions:** Thank you for participating in our study. Please answer all questions as best you can. If you have any questions while completing this assessment, please ask the experimenter.

What does the CPU use ROM for when executing programs?

a. Short term storage
**b. To hold instructions necessary to reboot**
c. Long term storage
d. To bypass the operating system

What happens to RAM when an application program terminates?

a. It lets the user know the program has terminated
b. Nothing happens to RAM
c. **The programs contents disappear**
d. The operating system goes to the hard disk and ROM

How does the CPU know when you give a print command?

a. The command goes from the monitor to the operating system to the CPU
b. The command goes from the monitor to RAM to the CPU
c. The command goes to the printer then the CPU
**d. The command goes to the CPU then the printer**

How does the CPU communicate with peripherals?

a. Through the operating system and RAM
b. Through RAM only
c. Through the operating system only
**d. Through the CPU only**

How does the computer access data from a magnetic tape cartridge?

a. The same way it does from a floppy disk
**b. The same way it does from cassette**
c. The same way it does from a compact disk
d. The same way it does from a DVD

How does the computer locate information stored on a floppy disk?

a. It searches very quickly from the beginning to the location
b. It searches very quickly from the last place data was stored
**c. It searches very quickly directly to the location**
d. It searches very quickly from the beginning of the disk

What does it mean to say the Internet is a society?

a. Because it is global, with no national boundaries or borders
**b. Because it has its own rules**
c. Because it is the only thing in the world that has no rules: it is anarchy
d. Because much of what goes over the Internet involves social communication

If you wanted to communicate emotion over the Internet, how should you do it?

**a. Your should use various symbols**
b. You would do so the same way you would in any letter sent by regular mail
c. You should always keep e-mail messages short and to the point
d. The rules of netiquette say keep e-mail short and to the point

How can people traveling in China check their e-mail sent to their computer back in Memphis?

a. Use the FTP protocol to do remote log on to the Internet from China

b. Use a LAN protocol to do remote log on to the Internet from China
**c. Use the Telnet protocol to do remote log on to the Internet from China**
d. Use the SLIP protocol to do remote log on to the Internet from China

How is it possible to download a computer game onto your own computer?

a. Through the LAN protocol
b. Through the Telnet protocol
c. Through the SLIP protocol
**d. Through the FTP protocol**

When you use terminal emulation, what is required to communicate over the Internet?

a. You need a SLIP connection
b. Your need a direct connection, like Road Runner
**c. Your need a standard phone line**
d. You need an Internet 2 connection

If you want to connect to the Internet using a gateway connection, what else do you need?

a. You need a direct connection and a network card
**b. You need a modem and an on-line service host**
c. You need a LAN protocol and a network card
d. You need a modem, a LAN protocol, and a network card

If several of your files are suddenly damaged, what is probably happened?

a. The CPU is faulty and you need to replace the motherboard
b. The operating system has malfunctioned
c. The hard drive has malfunctioned
**d. A utility program has malfunctioned**

If you are having problems copying files from one storage device to another, what is most likely responsible?

a. Your BIOS have malfunctioned
b. Your hard drive is damaged
**c. A utility program has malfunctioned**
d. You need to reinstall your operating system

Why is only part of the operating system stored in ROM on modern laptop computers?

a. ROM is not as fast as RAM
**b. It is difficult to upgrade**
c. It does not have the capacity to hold recent operating systems
d. ROM is difficult to access relative to the hard drive

When you first turn on your computer, what happens?

**a. It loads the operating system**
b. It sends a message to ROM and RAM
c. IT sends a message to the hard disk and RAM

d. IT sendz a message to the CPU and RAM

When a person is using an application program like a word processor, where is it in relation to the person, RAM, ROM, the hard drive, and the interface?

a. Between the operating system and RAM
**b. Between the interface and RAM**
c. Between the person and RAM
d. Between the person and the interface

When a person is using a computer, where is the operating system in relation to the person, running application, the interface, and the hardware?

a. It is between the interface and the running program
**b. It is between the application and the computer hardware**
c. It is between the interface and the computer hardware
d. It is between the person and the application

## Appendix B: Motivated Strategies Learning Questionnaire

### Part A. Motivation

The following questions ask about your motivation for and attitudes about your classes. **Remember there are no right or wrong answers, just answer as accurately as possible.** Use the scale below to answer the questions. If you think the statement is very true of you, circle 7; if a statement is not at all true of you, circle 1. If the statement is more or less true of you, find the number between 1 and 7 that best describes you

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Not at all true of me | | | | | | Very true of me |

1. I prefer course material that really challenges me so I can learn new things.
2. If I study in appropriate ways, then I will be able to learn the material in my courses.
3. When I take a test I think about how poorly I am doing compared with other students.
4. I think I will be able to use what I learn in this study in other courses.
5. I believe I will receive an excellent grade in my classes.
6. I'm certain I can understand the most difficult material presented in my classes.
7. Getting a good grade in my classes is the most satisfying thing for me right now.
8. When I take a test I think about items on other parts of the test that I can't answer.
9. It is my own fault if I don't learn the material in my courses.
10. It is important for me to learn the course material in my classes.
11. The most important thing for me right now is improving my overall grade point average, so my main concern in my classes is getting a good grade.
12. I'm confident I can learn the basic concepts taught in my courses.
13. If I can, I want to get better grades in my classes than most of the other students.
14. When I take tests I think of the consequences of failing.
15. I'm confident I can understand the most complex material presented by the instructor in my courses.
16. In my classes, I prefer course material that arouses my curiosity, even if it is difficulty to learn.
17. I am very interested in the content area of my courses.
18. If I try hard enough, then I will understand the course material.
19. I have an uneasy, upset feeling when I take an exam.
20. I'm confident I can do an excellent job on the assignments and tests in my courses.
21. I expect to do well in my classes.
22. The most satisfying thing for me in my courses is trying to understand the content as thoroughly as possible.
23. I think the course material in my classes is useful for me to learn.
24. When I have an opportunity in my classes, I choose course assignments that I can learn from even if they don't guarantee a good grade.
25. If I don't understand the course material, it is because I didn't try hard enough.
26. I like the subject matter of my courses.
27. Understanding the subject matter in my courses is very important to me.
28. I feel my heart beating fast when I take an exam.

29. I'm certain I can master the skills being taught in my classes.
30. I want to do well in my classes because it is important to show my ability to my family, friends, employer, and others.
31. Considering the difficulty of my courses, the teachers, and my skills, I think I will do well in my classes.

### Part B. Learning Strategies

The following questions ask about your motivation for and attitudes about your classes. **Remember there are no right or wrong answers, just answer as accurately as possible.** Use the scale below to answer the questions. If you think the statement is very true of you, circle 7; if a statement is not at all true of you, circle 1. If the statement is more or less true of you, find the number between 1 and 7 that best describes you

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Not at all true of me | | | | | | Very true of me |

32. When I study the readings for my courses, I outline the material to help me organize my thoughts.
33. During class time I often miss important points because I'm thinking of other things.
34. When studying for my courses, I often try to explain the material to a classmate or friend.
35. I usually study in a place where I can concentrate on my course work.
36. When reading for my courses, I make up questions to help focus my reading.
37. I often feel so lazy or bored when I study for my classes that I quit before I finish what I planned to do.
38. I often find myself questioning things I hear or read in my courses to decide if I find them convincing.
39. When I study for my classes, I practice saying the material to myself over and over.
40. Even if I have trouble learning the material in my classes, I try to do the work on my own, without help from anyone.
41. When I become confused about something I'm reading for my classes, I go back and try to figure it out.
42. When I study for my courses, I go through the readings and my class notes and try to find the most important ideas.
43. I make good use of my study time for my courses.
44. If course readings are difficult to understand, I change the way I read the material.
45. I try to work with other students from my classes to complete the course assignments.
46. When studying for my courses, I read my class notes and the course readings over and over again.
47. When a theory, interpretation, or conclusion is presented in my classes or in the readings, I try to decide if thee is good supporting evidence.
48. I work hard to do well in my classes even if I don't like what we are doing.
49. I make simple charts, diagrams, or tables to help me organize course material.
50. When studying for my courses, I often set aside time to discuss course material with a group of students from the class.

51. I treat the course material as a starting point and try to develop my own ideas about it.
52. I find it hard to stick to a study schedule.
53. When I study for my classes, I pull together information from different sources, such as lectures, readings, and discussions
54. Before I study new course material thoroughly, I often skim it to see how it is organized.
55. I ask myself questions to make sure I understand the material I have been studying in my classes.
56. I try to change the way I study in order to fit the course requirements and the instructor's teaching style.
57. I often find that I have been reading for my classes but don't know what it was all about.
58. I ask the instructor to clarify concepts I don't understand well.
59. I memorize key words to remind me of important concepts in my classes.
60. When course work is difficult, I either give up or only study the easy parts.
61. I try to think through a topic and decide what I am supposed to learn from it rather than just reading it over when studying for my courses.
62. I try to relate ideas in one subject to those in other courses whenever possible.
63. When I study for my courses, I go over my class notes and make an outline of important concepts.
64. When reading for my classes, I try to relate the material to what I already know.
65. I have a regular place set aside for studying.
66. I try to play around with ideas of my own related to what I am learning in my courses.
67. When studying for my courses, I write brief summaries of the main ideas from the readings and my class notes.
68. When I can't understand the material in my courses, I ask another student in my class for help.
69. I try to understand the material in my classes by making connections between the readings and the concepts from the lectures.
70. I make sure that I keep up with the weekly readings and assignments for my courses.
71. Whenever I read or hear an assertion or conclusion in my classes, I think about possible alternatives.
72. I make lists of important items for my courses and memorize the lists.
73. I attend my classes regularly.
74. Even when course materials are dull and uninteresting, I manage to keep working until I finish.
75. I try to identify students in my classes whom I can ask for help if necessary.
76. When studying for my courses I try to determine which concepts I don't understand well.
77. I often find that I don't spend very much time on my courses because of other activities.
78. When studying for my classes, I set goals for myself in order to direct my activities in each study period.
79. If I get confused taking notes in class, I make sure I sort it out afterwards.
80. I rarely find time to review my notes or readings before an exam.

81. I try to apply ideas from course readings in other class activities such as lecture and discussion.

**Appendix C: Agent Persona Questionnaire**

Please respond to the following statements and questions about your most recent interaction with AutoTutor.

**I enjoyed interacting with AutoTutor during my most recent session.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I felt that my interaction with AutoTutor during my most recent session was comparable to an interaction with a human tutor.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I felt engaged during the most recent tutoring session.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I felt that AutoTutor was difficult to use and work with during my most recent session.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I felt that I learned new information from AutoTutor during my most recent session.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**Understanding the material in my most recent session was important to me.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**While I was covering the material in my most recent session, I tried to make everything fit together.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I felt that AutoTutor was knowledgeable during my most recent session.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**The feedback from AutoTutor during my most recent session was appropriate with respect to my progress.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

## Appendix D: Confederate Perception Questionnaire

Please respond to the following statements and questions about your interaction with the confederate.

**I found the confederate to be believable.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I found the confederate to be attractive.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I found the confederate to be likeable.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I felt that that the confederate was agreeable.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I felt that the confederate was able to answer my questions during the study.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**I felt that the confederate was knowledgeable.**

| Strongly Disagree | Disagree | Somewhat Disagree | Somewhat Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |

**Appendix E: Graesser and Person Question Taxonomy**

*LEVEL 1: SIMPLE or SHALLOW*

| | |
|---|---|
| 1. Verification | Is X true or false? Did an event occur? |
| 2. Disjunctive | Is X, Y, or Z the case? |
| 3. Concept completion | Who? What? When? Where? |
| 4. Example | What is an example or instance of a category? |

*LEVEL 2: INTERMEDIATE*

| | |
|---|---|
| 5. Feature specification | What qualitative properties does entity X have? |
| 6. Quantification | What is the value of a quantitative variable? How much? |
| 7. Definition questions | What does X mean? |
| 8. Comparison | How is X similar to Y? How is X different from Y? |

*LEVEL 3: COMPLEX or DEEP*

| | |
|---|---|
| 9. Interpretation | What concept/claim can be inferred from a pattern of data? |
| 10. Causal antecedent | Why did an event occur? |
| 11. Causal consequence | What are the consequences of an event or state? |
| 12. Goal orientation | What are the motives or goals behind an agent's action? |
| 13. Instrumental/procedural | What plan or instrument allows an agent to accomplish a goal? |
| 14. Enablement | What object or resource allows an agent to accomplish a goal? |
| 15. Expectation | Why did some expected event not occur? |
| 16. Judgmental | What value does the answerer place on an idea or advice? |