

University of Memphis

University of Memphis Digital Commons

Electronic Theses and Dissertations

11-28-2012

Positive Versus Negative Agents: The Effects of Emotions on Learning

Carolyn McGregor Forsyth

Follow this and additional works at: <https://digitalcommons.memphis.edu/etd>

Recommended Citation

Forsyth, Carolyn McGregor, "Positive Versus Negative Agents: The Effects of Emotions on Learning" (2012). *Electronic Theses and Dissertations*. 606.
<https://digitalcommons.memphis.edu/etd/606>

This Thesis is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khhgerty@memphis.edu.

POSITIVE VERSUS NEGATIVE AGENTS: THE EFFECTS OF EMOTIONS ON
LEARNING

by

Carolyn McGregor Forsyth

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Major: Psychology

The University of Memphis

December 2012

ACKNOWLEDGEMENTS

This research was supported by the Institute for Education Sciences, U.S. Department of Education, through Grant R305B070349 to Northern Illinois University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Abstract

Forsyth, Carolyn McGregor. MS. The University of Memphis. December 2012. Positive Versus Negative Agents: The Effects of Emotions on Learning. Major Professor: Arthur Graesser, Ph.D.

The current study investigates the impact of affect, mood contagion, and linguistic alignment on learning during tutorial conversations between a human student and two artificial pedagogical agents. The study uses an Intelligent Tutoring System known as *Operation ARIES!* to engage students in tutorial conversations with animated agents. In this investigation, 48 college students ($N = 48$) conversed with pedagogical agents as they displayed 3 different moods (i.e. positive, negative, and neutral) along with a control condition in a within-subjects design. Results indicate that the mood of the agent did not significantly impact student learning even though mood contagion did occur between the artificial agent and the human student. Learning was influenced by the student's self-reported arousal level and the alignment scores that reflected a shared mental representation between the human student and the artificial agents. The results suggest that arousal and linguistic alignment during tutorial conversations may play a role in learning.

TABLE OF CONTENTS

Section	Page	
1	Introduction	1
	Agent-based Intelligent Tutoring Systems	3
	Agent Emotions and Learning	4
	Learner Emotions and Learning	5
	Mood Contagion	10
	Linguistic Alignment	13
	Purpose of the Experiment	16
2	Method	17
	Participants	17
	Design	17
	Materials	18
	Agent-human interaction	18
	Content covered in learning session	21
	Manipulation check	22
	Assessment of learning gains	22
3	Procedure	23
4	Measurements	24
5	Analyses	28
	Linguistic Measures	33
	Coh-Metrix measures	33
	LIWC Measures	34
6	Discussion	35
	References	39
	Appendices	45
	A. Example Script	45
	B. Example Conditions	46
	C. Counter-balanced Conditions	48
	D. Example Questions for Pre-Test and Post-Test	49

Introduction

The mood of teachers likely has an impact on the learning gains of their student. For example, one could imagine that an upbeat high-school teacher's students might learn more than a frustrated teacher's students because upbeat teachers may create an atmosphere of positivity, hopefulness, and adventure. Frustrated teachers, on the other hand, may create an environment of negativity, hopelessness, and disapproval. Some research projects have already demonstrated that learners' emotions can have serious consequences on their learning gains (Pekrun, 2006; Pekrun, Maier, & Elliot, 2009; Zeider, 2007), and that teachers' emotions can significantly impact students emotions (Sullins, Craig, & Graesser, 2009). However, the research is contradictory and it is important to uncover which teacher emotions lead to increased learning, which lead to decreased learning, versus which emotions have no impact on learning. Making this distinction could aid in the development of better teacher training programs and the design of intelligent tutoring systems that are more sensitive to learners' emotional needs.

The study of emotions encompasses a large arena of affective states. In order to distinguish between these states, affect can be grouped into specific emotions and moods (D'Mello & Graesser, in press). In day-to-day experiences, discrete emotions, which arise in response to natural circumstances, are usually transient (or short-lived), but they sometimes may affect the overall mood of a person for a few days. For example, a teacher who is involved in a fender-bender on the way to work may be intensely frustrated at the time of the accident. Eventually, as the day passes, that intense emotion may subside. However, it may translate to a generalized negative mood which may last until the car is out of the shop.

Both emotions and moods are often categorized by valence, either positive or negative (Eisenhower, Frank, & Carello, 2010; Russell, 2003). By categorizing emotions on these two polarities, it allows for global investigations into the nature of emotions. Contradictory evidence exists as to whether positively- or negatively-valenced emotion as displayed by the teacher are more supportive of student's learning (Sullins et al., 2009; Isen, 2008). Surely, the student's affective response and mutual understanding with the teacher will moderate these effects. This thesis will focus on the impact of the mood of the teacher on the learning gains of a student. It will also explore possible moderating factors, namely as linguistic alignment and mood contagion, as described below.

This investigation into mood, learning, linguistic alignment, and mood contagion will employ an artificial environment with animated pedagogical agents. These environments allow the pedagogical agents (i.e., talking heads) to exhibit a consistent display of one affective state, without the variability of emotions that arise in human teachers. Many studies (Craig, Graesser, Sullins, & Gholson, 2004; D'Mello, Picard, & Graesser, 2007; Sullins et al., 2009) have used artificial environments with agents in empirical investigations into the relationship between mood and learning gains. The current study will similarly use an agent-based intelligent tutoring system (ITS) that helps students learn by holding a conversation in natural language.

The next section will review agent-based ITSs, followed by a discussion of empirical and theoretical research on emotions and learning. The section will subsequently discuss insights from research on linguistic alignment and mood contagion during learning. With this theoretical and empirical context developed, the thesis will describe how insights from these areas motivated the design of the experiment.

Agent-based Intelligent Tutoring Systems

ITSs have been designed to instruct students on various topics and have been found to foster learning gains (Graesser, Conley, & Olney, 2012). Agent-based systems are helpful in aiding students in the learning experience by engaging students in an informal tutorial conversation (Graesser, Jeon, & Dufty, 2008; Reeves & Nass, 1996) as well as providing a multi-media interactive environments (Moreno & Mayer, 2002). In particular, *Operation ARIES! (Acquiring Research Investigative and Evaluative Skills;* Millis et al., 2011; Halpern et al., in press) is an agent-based ITS designed to teach the scientific method, such as the definition of dependent variables, replication, and subject bias. Two pedagogical agents, both a teacher and student agent, are used to facilitate natural language dialogues (conversations between two animated agents and one human student). *Operation ARIES!* adopts an e-Text, multiple choice questions, and three-way conversations in a game environment in order to facilitate learning.

Another example of an agent-based ITS is AutoTutor which instructs students on physics (VanLehn et al., 2007) and computer literacy (Graesser et al., 2004) through natural language dialogue. AutoTutor has yielded sizeable learning gains on par with effect sizes found in one-on-one human tutoring (Graesser, Chipman, Haynes, & Olney, 2005; Graesser et al., 2004; VanLehn et al., 2007).

Describing all the various agent-based ITSs is beyond the scope of this paper, but a few additional examples are noteworthy. There are agent-based systems that instruct students on how to regulate learning (Azevedo, 2007), ask deep questions (Gholson & Craig, 2006), read at deeper levels of comprehension (McNamara, Levinstein, &

Boonthum, 2004), and master geometry (Burlison, 2006). Other ITS's incorporate emotions, which is the focus of this thesis.

Agent Emotions and Learning

Sullins et al. (2009) investigated the effects of the mood of animated agents on the learning gains of students in the domain of computer literacy. The participants were divided into a two conditions (positive affect or negative affect). The authors created a presentation of the material implemented in AutoTutor in order to teach the participants 12 subtopics in computer literacy. The face of the agent (embedded in the instruction) was altered in order to present the material in either a positive or negative valence. That is, each group was shown either a seemingly happy or a sad artificial agent at the beginning and end of each of the 12 sub-sections. Learning gains were calculated for the two separate conditions and the results were surprising. Students actually performed better in the negative affect condition.

The "illusion of knowing" hypothesis (Glenberg & Epstein, 1985) is a possible explanation for these results. The hypothesis posits that people believe that they know more than they actually do. In Glenberg and Epstein's experiment, students were asked to read a text and identify inconsistencies. Some students failed to notice blatant contradictions in the text. Interestingly, these same students gauged their own comprehension as very high. Baker (1985) suggests that we judge our understanding based on three specific levels of comprehension: lexical, syntactic, and semantic. Understanding the lexical level involves simply comprehending the words that are used. The syntactic level specifies the structure of the sentence, rather than just the individual words. Both the lexical and syntactic levels must be achieved for one to reach the

semantic level, or the comprehension of the actual meaning of the text. A flaw in judgment at any level will result in a final misunderstanding of the meaning. A difference may potentially occur between the reader's understanding and a deep understanding of the text (Glenberg & Epstein, 1985). This creates the gap between knowing material and the illusion of knowing.

As applied to the Sullins' et al. (2009) results, the illusion of knowing hypothesis explains positive affect as misleading the students that they understand the technical material and that attribution hinders learning gains. For example, if the agent appears to be happy when a false answer is presented, then the students might believe that the incorrect answer is in fact correct. The students believe that they know the material when in fact they do not, which decreases knowledge acquisition. The mood as information theory (Schwarz & Clore, 1983) also supports this claim. It states that expressions of moods can provide information about ambiguous situations. Students who interacted with a positive agent may have made the appraisal that the agent only is positive when they are performing well. Consequently, students decreased their effort and attentional allocation, which ultimately decreased learning.

Learner Emotions and Learning

Positive emotions in students may be more conducive to creative problem solving according to Isen, Daubman, and Nowicki (1987). Participants were divided into two conditions in the Isen et al. experiment. The participants in the experimental condition were induced to have a positive mood, while those in the control condition did not receive mood induction. Participants in both conditions worked on Duncker's (1945) candle task and the Remote Associates Test (RAT; Mednick, Mednick, & Mednick,

1964), which tap problem solving and creativity skills, respectively. The goal of Duncker's candle task is to attach the candle to a wall in a manner in which the lit candle will not spill wax on the table or the floor. The props provided include a book of matches, a candle, and a box of tacks. Solving this problem is no easy feat and requires thinking "out of the box". The control condition solved the problem 11-16% of the time. The experimental group, those with a positive affect, were able to correctly solve the problem 58-75%. Clearly, there was a dramatic increase of 47- 59% between the control and experimental groups. The second task given was the RAT. In this task, the subject is shown three words and asked to generate a fourth associated word. For example, the given words could be *mower*, *atomic*, and *foreign*. A word that is correctly associated with the other three could be *power*. The results were once again in favor of the positive condition, suggesting that positive affect increases both problem solving and creativity.

Positive affect has also resulted in greater flexibility of attention and cognitive processing (Isen, 2008). As participants completed a simple task, one group was given positive-feedback whereas the other group was given negative-feedback, an effective means of mood induction. After receiving the feedback, the researcher asked the participants about seemingly unrelated interactions between human confederates that had occurred nearby. Those who just received positive-feedback were able to give clearer and more accurate descriptions than those who received negative-feedback. These results can be interpreted to be support of the hypothesis that positive affect facilitates cognitive flexibility.

In another study, it is suggested that positive affect increases global information-processing (Glasper & Clore, 2002). In the experiment, Glasper and Clore induced two

groups of participants to have either a positive or negative mood. After induction, both groups were momentarily presented with a structural picture. After removing the picture, the subjects were asked to iteratively draw the same picture 10 times. The positively-induced participants were better able to replicate the original picture than the negatively-induced subjects. Furthermore, the replications created by the positive affect condition maintained the original schema and title of the picture whereas the negative affect group did not. The authors hypothesized this to be a result of positive affect leading to a more global perspective.

In order to test this hypothesis, Glasper and Cloire (2002) performed a follow-up experiment. In the experiment, participants were induced into varying affective conditions (positive, negative or neutral), and were then presented with a complex geometric object. Next, the participants were asked to match the given shape with a new object. The comparisons could be based more off of local or global aspects within the original geometric object. For example, a global aspect might be a larger cube whereas the local shape might be a smaller rectangle within the cube. Participants in the positive affect condition were more likely to globally group shapes than participants in the negative affect condition, thus supporting the researchers' hypothesis.

Increased creative problem solving, cognitive flexibility and global information processing can be explained by a theory known as the "broaden and build theory" (Frederickson, 1998). Inspiration for the theory can be found from early human survival of the fittest. Evolutionary purposes would require one to have an immediate fight or flight reaction in the face of a threat, such as walking up on a bear. When such a threat was not present, the early humans would need to expand their resources in order to be

better prepared for future negative encounters. The *broaden and build theory* is based on the idea that negative emotions tend to require immediate reactions whereas positive emotions allow one to view the surroundings (broaden) and accrue more resources in order to increase the probability of survival (build). Basically, the positive affective states will augment the relationship between thought and action and thus build creative and investigative capabilities. According to this theory, student's experiencing a positive mood should be more capable of learning new material.

Another possibility is that valence itself does not affect the functioning of cognitive processes. The *arousal theory* suggests that only the intensity (not the valence) of an emotional trigger is important (Bradley, Codispoti, Cuthbert, & Lang, 2001; Lang, 1995). Schimmack (2005) investigated the relationship between attention and both affect valence and intensity using visual pictures of emotional situations as stimuli. Findings from this experiment revealed that emotional valence of the pictures did not affect attention, but the intensity did have an effect. Other researchers predict that arousal is important for only simple tasks (Martindale, 1981; Yerkes & Dodson, 1908), but not for more intricate tasks such as creative problem-solving (Isen et al., 1987). Therefore, the empirical evidence for the effects of emotional stimuli on cognitive processes is still incomplete.

Pekrun and colleagues (2009) suggest a multi-layered model of learning, motivation, and emotion. This model, known as the control-value theory of achievement emotions, posits that motivation or goal-orientation of the student is correlated with specific discrete emotions such as enthusiasm and boredom during learning. According to this theory, learners' goal-orientation can be focused on either *mastery* or *performance*. A

learner with mastery goal-orientation wants to learn the material for the sake of acquiring knowledge. On the contrary, a student who is performance goal-oriented wants to only learn enough information to perform well on an exam. The difference between these two types of learners is that the first is more likely to achieve deep, meaningful learning, while the second is likely to only acquire enough shallow knowledge to get a passing grade in fear of failure or reprimand (Pekrun, 2006). Resulting emotions derived from these attributions can positively or negatively affect learning gains. An example of a resulting emotion could be the happiness experienced by the mastery-oriented student after successfully comprehending a difficult topic. As predicted by the model, both the student's goal-orientation and emotional state predict learning gains in an undergraduate classroom. Specifically, a mastery-goal oriented student experiencing enthusiasm has higher learning gains than a student experiencing boredom that results from lack of interest or motivation (Pekrun et al., 2009). In summary, the high-level goal orientation associates positive emotions towards approaching the task which leads to a high level of motivation and allows for flexible learning strategies and ease of accessing cognitive resources (Linnenbrink & Pintrich, 2002; Pekrun et al., 2009).

Contradictions exist as to whether positively-or negatively-valenced moods and emotions are more likely to yield increased learning gains. Teachers believe that they are more effective when in positive mood (Franzel, Goetz, Ludtke, Pekrun, & Sutton, 2009). This is understandable in the sense that enthusiasm has been shown to increase the teachers' effectiveness (Klussman, Kunter, Trautwein, Ludtke, & Baumert, 2008) by increasing the efficiency of the instruction (Franzel et al., 2009). However, Sullins and colleagues (2009) found higher learning gains when an artificial agent embedded in the

lesson displayed a negative affective state. With conflicting theories and research, it is important to explore alternate explanations for differential learning gains as a function of the pedagogical agent's valence. Possible moderating factors include both the student's affective response to the teacher's mood as well as the properties (e.g., student engagement and attainment of common ground) of the conversation between themselves and the artificial agent. The nature of these two aspects may be captured by mood contagion and linguistic alignment. Though the empirical evidence and theories provided about the phenomena are based on human-to human interaction, the findings resulting from such studies may also apply to human-agent interaction. The reason is that some evidence exists suggesting that human's perceive interactions with agents similarly to human-to human interaction (Clark & Mayer, 2002; Reeves & Nass, 1996). With this in mind, we now turn to the literature on mood contagion and linguistic alignment.

Mood Contagion

According to Hsee, Hatfield, and Carlson (1990), emotional contagion is defined as "the tendency to mimic the verbal, physiological, and or behavioral aspects of another person's emotional experience/expression and thus to experience/express the same emotions oneself". Emotional contagion has been reported anecdotally throughout history in order to explain "mob mentality". This phenomenon refers to a group of people experiencing the same passionate emotions. Mob mentality has been reported in both religious and political contexts (Hatfield & Rapson, 2004). Early clinicians such as Jung (1968) and Reik (1948) believed strongly in emotional contagion. Jung and Reich claimed that they could understand their patients' emotions by monitoring their own feelings. For example, if the clinician noticed that he was feeling anxious, he operated

under the assumption that the patient was experiencing anxiety. Modern day clinicians have also reported experiencing emotional contagion (Howes, Hokanson, & Lowenstein, 1985). The phenomenon has not only been reported by clinicians, but also social psychologists, animal researchers, and historians (Hatfield, Cacioppo, & Rapson, 1993).

Beyond anecdotal evidence, empirical evidence has shown both the existence and circumstances of emotional contagion. In a relevant experiment, a student–teacher environment was manipulated in order to study mood contagion and dyadic interactions (Hsee et al., 1990). The researchers created high-stakes environments where the pressure and motivation were increased. Specifically, the “teacher” participants were ostensibly told they were administering electric shocks to the students who gave incorrect answers. In this circumstance, the affective state of the more submissive partner, the student, was more likely to be acquired by the more powerful partner, the teacher. Additionally, in the study of group dynamics, correlations have been found between the affective state of the leader and other members of the group (Sy, Côté, & Saavedra, 2005).

Bavelas and colleagues (1986) propose that expression of another’s affective state or “vicarious emotion” is explained by a *communication hypothesis*, which postulates that physical expressions are not intra-personal responses but rather social responses. For example, a woman falls and hurts her ankle and a passerby turns down his lips to form a frown. The man frowns because it is the appropriate social response. He does not necessarily frown because he is truly feeling sad because the woman fell. She supports this model with research suggesting that participants show more facial and physically “empathetic” responses when the sender (the one originally experiencing the emotion) is

directly in front of them and making eye-contact (Bavales, Black, Lemery, & Mullett, 1986).

Hatfield and colleagues (1993) specifically describe emotional contagion to be manifested by the sender's behaviors, facial expressions, and postural movements. They postulate that this synchrony occurs to create a real-time constant feedback mechanism enabling non-explicit information to be transmitted to the receiver (the person who is mimicking the sender). In practical terms, the receiver is unconsciously mimicking movements, facial expressions and behaviors of the sender and then subsequently applying his own emotions to consciously gauge the emotions of the sender (Hatfield et al., 1993).

Other research supports Hatfield and colleagues conclusions (Duclos et al., 1989; Neumann & Strack, 2000). Neumann and Strack (2000) also agree that mood contagion is automatic, non-conscious, and transmitted via behavior. First the physical attributes are mimicked, and then the mind deciphers the mood state based entirely off of the activation in the autonomic nervous system. Therefore, mood contagion must occur only by the emotional expression of the sender's behavior (both verbal and non-verbal). Clearly, affective alignment or mood contagion may be achieved through multiple channels including physical as well as linguistic. In the current study, the focus is on emotions during a tutorial conversation between a human and an artificial agent (talking head). Therefore, the similarity of other features such as linguistic alignment during communication may shed additional light on the intricate nature of emotions and learning.

Linguistic Alignment

Previous research has suggested that synchrony often exists during communication between two humans. One such study, *The Chameleon Effect* (Chartrand & Bargh, 1999), describes human mimicry in nonverbal behaviors, accents, and speech patterns. Likewise, when communicating with others, humans will adapt themselves to the structure of politeness demonstrated by the dominant party in the conversation through linguistic alignment. This alignment has been attributed to a desire to "save face," and has been demonstrated within and between different cultures. The alignment could be reflective of a type of genetic coding in humans. It is theorized that evolution has adapted genetic coding in humans to allow for mimicry because aligning with peers may increase the probability of one being accepted in society (Brown & Levinson, 1978). Another evolutionary purpose of human-to-human mimicry may be to acquire new traits necessary for survival of an individual or a species (Lakin, Jefferis, Cheng, & Chartrand, 2003).

Many other possible reasons for mimicry have been proposed. One study suggests that there is a higher lexical correlation between people who trust each other more (Scissors, Gill, & Gergle, 2008). This finding demonstrates that perhaps we do not blindly align with our conversational partners. Instead, our appraisal of the person with whom we converse, or our relationship with that person, can affect our conversational alignment. Another possibility is that people tend to mimic others if there are similarities or attraction between both parties; this is referred to as the *similarity-liking hypothesis* (Byrne, 1971). This hypothesis claims that we align more with conversational partners who are similar to ourselves, as opposed to those with whom we perceive to have no

common ground. Further research has indicated that mimicry occurs at greater levels with higher levels of attraction (Bernieri, Reznick, & Rosenthal, 1988). The *coordination-rapport hypothesis* (Tickle-Degnen & Rosenthal, 1987) posits that this phenomenon could occur as a means of facilitating both conversation and the building of a relationship between the two people. Furthermore, the *communication-accommodation theory* (Giles et al., 2005) suggests that mimicry occurs in order to facilitate communication and social acceptance. In order to maintain personal identity, the alignment only occurs at the necessary level needed to reach the communicative goal. Thus, a person may only align with another person to the extent that the conversational goals can be met without sacrificing one's perception of him or herself. Finally, the *common ground hypothesis* (Clark & Schaeffer, 1987) states that synchrony occurs when the two people have yet to reach a common ground in regards to beliefs or opinion. However, once this common ground is reached, Clark and Schaeffer (1987) state that the synchrony (or alignment) will slowly fade, as it has already served its purpose. Common ground may be especially important in the learning environment as the goal is to create a shared understanding between a teacher and a student.

Another explanation is the *coordination-engagement hypothesis* (Niederhoffer & Pennebaker, 2002). The hypothesis proposes that alignment does not depend on whether or not the person wants to be liked or accepted, but rather is solely a function of engagement. For example, if the more dominant person is angry, then the submissive partner will also display this emotion if sufficiently engaged in the conversation. The hypothesis emerged from a study finding that synchrony existed between dyads in a computer-mediated environment on multiple levels, including linguistic, affective and

cognitive categories all measured by the textual input of the user. This research contradicts the *coordination-rapport hypothesis* (Tickle-Degnen & Rosenthal, 1987) because the submissive partner does not conform to a positive affective linguistic style in order to please the dominant partner. Though the exact nature of some of these hypotheses will not be tested in the current investigation of affect, alignment and learning, they provide valuable reasons for why mimicry occurs and could be adopted to direct future studies.

Linguistic alignment is widespread throughout society. It occurs in both computer-mediated discourse as well as in conversations occurring in natural settings (Mehi & Pennebaker, 2003). The phenomenon even permeates the cultural boundaries of conversational participants (Cassell & Tversky, 2005). With such a pervasive occurrence of alignment, it seems probable that the construct will also be found in agent to human interaction in an intelligent tutoring system. As previously stated, humans perceive conversations with artificial agents to meet the standards of real social interactions within a learning environment (Clark & Mayer, 2002; Reeves & Nass, 1996). In one experiment, the Reeves and Nass reported that humans were less likely to give negative feedback about an artificial agent on the same computer in which they interacted with the agent. This is not to say that they believe the agents are human, but rather that the conversation is authentic to the point that the agent should be treated in line with social conversational rules. For example, when one agent is speaking, the other should be listening (Clark & Mayer, 2002). With these similarities between human-to-human vs. human-to-agent interactions, it is possible that the alignment phenomena of mood contagion and linguistic alignment both occur and impact interactions with artificial agents in an ITS.

Purpose of the Experiment

The current investigation into mood, learning, linguistic alignment, and mood contagion was conducted in the context of an agent-based ITS. Alterations in mood may be manifested in tone (Neumann & Strack, 2000), facial expression and bodily movements (D’Mello & Graesser, 2010), all of which can be generated by agents. In an agent-based system, “talking heads” speak to the student and various dimensions of the communication can be controlled. The tone was controlled by using an automated speech engineering system. Facial expressions were controlled by using the same agents for all participants. The body movements were virtually non-existent in the agents in the system. Both the mood and appearance of the animated agents were easily controlled and modified, unlike with a human confederate.

The purpose of the thesis is to primarily test two hypotheses that make predictions on the impact of emotional mood of an agent on student learning. A *positive mood facilitation* hypothesis predicts that a positive mood of the pedagogical agent will increase the student’s learning gains, relative to negative or neutral mood. In contrast, a *negative mood facilitation hypothesis* makes the opposite prediction. If one of these two hypotheses is confirmed, then there is the question of the extent to which learning is predicted by linguistic alignment and mood contagion. However, if neither are confirmed, additional questions remain. Specifically, it is necessary to discover whether or not the manipulation of the tutorial condition created comparable affective states reported by the participants. If the manipulation is successful, then the relationship between the reported affect of the student and learning can be investigated. Regardless of the success of this manipulation, the relationship between the linguistic measures of affect and alignment

will be compared to learning. Finally, the linguistic affective scores will be compared to the self-reported affective state of the student.

Method

Participants

The study included 48 (N = 48) undergraduate and graduate students at The University of Memphis. Out of the total number of participants, 4 were graduate students and the other 44 were undergraduate students. Subjects were recruited using Sona Systems, the University subject pool, as well as through flyers and word of mouth. Two types of compensation were offered: A monetary reward (\$15 for completion of the entire experiment) or course credit (2 hours course credit towards an Introduction to Psychology course). Only 13% of the students requested the monetary compensation (\$15), whereas the others preferred 2 hours of credit towards an Introduction to Psychology course.

Design

The two primary hypotheses were tested using a within-subjects design and a counter-balanced pretest and posttest. Participants interacted with an ITS by reading an E-book, answering multiple-choice questions and conversing in natural language dialog about four topics on the scientific method (see the following section for a detailed description of the ITS). A control chapter was in a “text only” format that had no dialog. The assignment of condition with positive, negative, neutral or control were counter-balanced across participants. However, the order of the topics was kept constant across all 4 conditions. Thus, each participant received each mood condition for one of the topics (counterbalanced across participants and topics) and one chapter served as a control (i.e., text only). In order to assess whether or not the manipulation indeed

induced the designated emotion, students were asked to report emotions 5 times throughout the experiment. The first reporting took place before interaction in order to gain a baseline, followed by sequential reports after each of the 4 topics within the system. Learning gains were assessed using a pretest and posttest consisting of multiple-choice and open-ended questions. Mood contagion and linguistic alignment were assessed by analyzing the linguistic input of the student during the tutorial conversations.

Materials

Agent-human interaction. *Operation ARIES!* (*Acquiring Research Investigative and Evaluative Skills*; Millis et al., 2011; Halpern et al., in press) is an adaptive ITS that teaches the scientific method through natural language conversations. After reading chapters in an eBook, *Operation ARIES!* tests the students with multiple choice questions to enable the system to adaptively place the student into one of three pedagogical modes. Students can be classified into low, medium, and high mastery on the basis of these tests. Two agents, both a teacher and student agent, are needed to exhibit the varying modes, as described below.

The three modes consist of vicarious learning, regular tutoring, and teachable agent. In the vicarious learning mode, the student simply watches the teacher agent teach the student agent, but periodically the human is asked a YES/NO question to make sure they are engaged. At the other extreme, in the teachable agent mode, the student teaches the student agent. For the purposes of this experiment, however, only the intermediate mode was used. This mode is referred to as regular tutoring and consists of the pedagogical agents asking the student a specific question about the current topic and scaffolding the students to help them articulate a pre-determined ideal answer.

The artificial agents scaffold the students by providing appropriate feedback, hints, prompts, and correcting misconceptions. For example, if students are unable to provide much information, they will first be given a hint. This is a broader statement which should help the student recall some information about the topic. If the student is still unable to provide a correct answer, the agents will give the student a prompt. A prompt should be easier to answer than a hint and only requires a single word or phrase as an answer. For example, a prompt might be, “It is not increase but what?” An example conversation can be found in Appendix A.

Every question asked by the pedagogical agents has a corresponding ideal answer. The typical ideal answer consists of three to four important words. For example, a question requiring such an answer could be, “Why are operational definitions important?”. The ideal answer is “Operational definitions are important so that other researchers can understand the variables in a study so that they can perform the study for themselves and get the same results.” The important words from this ideal answer are “understand”, “variables”, “study”, “same”, and “results”. Throughout the conversation with the agents, the artificial pedagogical agents attempt to get the student to articulate these five words. Semantic matches are deciphered by comparing the human input to the regular expressions (Jursfsky & Martin, 2008) that are implemented in *Operation ARIES!*. Regular expressions allow for more alternative articulations of an ideal answer. In the above example, the student needs to type “same” and “results” in the same utterance. The regular expression “(same|similar|identical).*result” constrains the matching by forcing the student to type two words in the same utterance. It allows for more numerous matches by allowing for synonyms of “same” to be accepted. In order to

accept an expectation as being covered, the student must reach a .51 threshold of overlap between the student's language over multiple conversational turns and a sentence-like expectation. In the above example, the student will have to say three of the five words in order to move forward to the next topic section.

For the purposes of this experiment, the mood of the agent was altered by changing the curriculum scripts, which is the pre-determined speech of the two pedagogical agents. An example of an altered script can be found in Appendix B. Both the teacher agent and student agent were in the same mood at the same time. So, during a "positive" chapter within the learning session, both the teacher and student agent exhibited a positive mood. During a "negative" chapter, both were in a negative mood. For each chapter, the moods remain either positive, negative, or neutral for both agents throughout an entire chapter.

The statements made by the agents were altered using the LIWC, an acronym for Linguistic Inquiry and Word Count, lexicon (Pennebaker, Francis, & Booth, 2001). This lexicon has numerous words associated with positive and negative affective states. For example, words with positive valence include "happy, curious, and awesome" whereas negatively-valenced words include "bored and sad". Therefore, in the positive condition of the given example the agent's input might be altered to "No. You are incorrect. Let's just go over the importance of these awesome things one more time. Why do we need to have operational definitions?" An example statement in the negative condition would be, "No. You are incorrect. Let's just go over the importance of these dull things one more time. Why do we need to have operational definitions?". As the reader may notice, the manipulation did not include a change in feedback. Therefore, if the student provided an

incorrect answer to a question, then the feedback was still “No, you are wrong.” The goal of the current study was to manipulate the mood of the teacher not the accuracy of feedback. Therefore, the emotive words were intended to be expressive of the teacher agent’s affective state independent of the student’s performance. An example of the three affective conversational conditions can be found in Appendix B.

The curriculum scripts covered three chapters of material with three separate chapters each designated to one mood condition (i.e., positive, negative, or neutral). The participants were exposed to a fourth chapter of content through the e-Text only. This chapter served as a control for the counter-balanced conversations with the artificial agents. In order to return the student to a baseline of emotion between the within-subjects conditions, each chapter began with the student answering 6 multiple-choice questions about the topic and reading a summarized chapter of the e-Book within ARIES.

Content covered in the learning session. All of the topics in *Operation ARIES!* are about Research Methods. Some of the information requires cumulative knowledge, meaning that is necessary for students to understand some topics in order to understand others. For example, students may not understand the concept of generalizing results if they have not been exposed to the meaning of replication. In the current study, this type of overlapping information might confound the results if one mood condition is more effective than another. Therefore, chapters that did not require much prerequisite knowledge dependent on a previously manipulated chapter were specifically chosen. In order to account for any necessary prior knowledge, a summary of needed information beyond the specific topic at hand was provided within the E-book chapter summary.

Manipulation check. In order to ensure that the manipulation of the tutorial conversation of the pedagogical agent is effective, students were asked to report their emotions on an affect grid (Russell, Weiss, & Mendelsohn, 1989). The grid is composed of two dimensions that include valence (the polarity of the emotion) and arousal (the intensity). The two constructs are plotted on a grid with the x axis representing valence and the y axis representing arousal. Students were given explicit instructions on how to complete these grids and a baseline state was recorded before interaction with the ITS.

Assessment of learning gains. Two similar, but not identical, versions of a learning gains assessment were created. These two versions are referred to as Version A and Version B. The learning gains assessments were counter-balanced, so both versions of the assessment were used as a pretest versus a posttest that assesses learning gains. The assessments were devised to measure both shallow and deep-level knowledge. According to Bloom's taxonomy (Bloom, 1956), a transition from recognition (multiple-choice questions) to recall (short answer) taps a deeper level of understanding. Therefore, the assessment questions progressively became more difficult as they transition from multiple-choice questions to short-answer. Due to time constraints within the experiment, only 1 of the 8 questions was short answer whereas the other 7 were multiple-choice.

The assessment questions did not only change in format but also by orientation. The questions gradually became isomorphic in order to tap a deeper level of knowledge. Isomorphic questions present the same material in a different context, which forces the student to transfer the original knowledge to a new domain or situation. For example, if the original question asks for a definition of a topic such as operational definitions, then the isomorphic question might ask for an operational definition for "people who read a lot

of books”. Appendix D provides example questions in both shallow and deep categories. Eight questions assessed the knowledge acquired for each chapter, which resulted in a total of 32 questions in Version A and 32 questions in Version B.

Procedure

Upon entering the lab, participants were given an Informed Consent form explaining what participation in the experiment will entail. Participants were then instructed that the purpose of the experiment was to investigate a new Intelligent Tutoring System and emotional responses. At the beginning of the experiment, participants were randomly assigned to different materials that varied: 2 tests (pre/post) x 4 moods (e.g., positive, negative, neutral, neutral_{text only}) in a counterbalancing scheme. As previously mentioned, two versions of the learning gains tests were created including a Version A and Version B. The assignment of the two tests was counter-balanced as pretest and posttest. Thus, within the four possible mood conditions (e.g., positive, negative, neutral, neutral_{text only}) and the counter-balanced learning gains assessments, there were 48 cells. Each subject was exposed to all three conversational mood conditions in a counterbalanced order as well as the text only condition. Within each of these groups, each participant was exposed to four topics of research methodology which were presented in a constant order across subjects. However, the actual topic order remained constant across subjects. A list of all of the possible conditions can be found in Appendix C.

After being randomly assigned to a specific group, each subject was first given a pretest (either Version A or Version B of the learning gains assessment) consisting of 32 questions. Next, instructions were given regarding completion of the affect grid. After

completing the standardized explanation of the grid, the researcher went to each participant individually to ensure understanding of the task by asking “How do you feel participating in this experiment?”. Students verbally replied and were asked to denote this emotion on the first affect grid. After completing the baseline affect grid, each participant interacted with *Operation ARIES!* in the respective assigned condition. During interaction with *Operation ARIES!*, students read an E-Text on each chapter as well as answered multiple-choice questions before engaging in a natural language dialogue with the two pedagogical agents. In accordance with the given condition, the participants conversed with the agents displaying the information in a specific mood, i.e. positive, negative, or neutral. After completion of each topic, participants were asked to fill out the subsequent affect grid. This process occurred iteratively across the chapters. In the text-only treatment, participants only read the E-Text and answered multiple-choice questions but did not participate in a tutorial conversation with the agents. Upon completion of the interaction with *Operation ARIES!*, the participant were given a 32-question posttest (either Version A or Version B of the learning gains assessment according to the counterbalanced order). Finally, participants were debriefed and thanked for their time.

Measurements

Learning gains were calculated based on pretest and posttest scores taking into account the varying levels of prior-knowledge of the students. As previously mentioned, two versions of the test exist. Both tests included 32 questions including both multiple-choice and open-ended questions. Both the pretest and the posttest were manually graded in accordance with the associated rubric. Then proportional learning gains (PLG) were

calculated using the formula $[(\text{posttest}-\text{pretest})/(1-\text{pretest})]$ (Jackson, Graesser, & McNamara, 2009).

However, upon initial inspection of the data, it became apparent that proportional learning gains may not be the best measure of learning due to the calculation on the item level. When negative learning gains are found within a specific topic using this formula, it is possible to occasionally reach a large negative value such as -4 which is not aligned with the PLG scale of (0-1). Because the experiment was counter-balanced with a participant in each of the 48 groups and learning gains scores (posttest-pretest) did not reveal such extreme outliers, the researcher made the decision to not remove participants. Therefore, in the following analyses assessing learning gains, posttest was used as the dependent variable and pretest was entered as a covariate.

The current experiment required measurements for mood contagion and linguistic alignment. In order to investigate mood contagion, the affective state of the student was assessed using Linguistic Inquiry and Word Count (LIWC; Pennebaker et al., 2001). LIWC is an automated system that provides frequency counts (in proportion to the total number) of words associated with different psychological categories (i.e., affective and cognitive) within a given text or discourse segment. Specifically, LIWC provides word frequency proportion scores for positive and negatively-valenced affective speech.

All of the student input was analyzed on a by-conversation basis using LIWC in order to gain word frequency proportion scores of positively and negatively-valenced affective words for each of the three conversational conditions (i.e., positive, negative,

and neutral). One measure of the mood of the student was represented by these proportion scores reported by LIWC for the positive, negative, and neutral conditions.

The second measure of mood was obtained from the self-reported levels of both valence and arousal on the affect grids. The investigation operates off of the assumption that students will return to a baseline of emotion after each interaction with ARIES. Therefore, the valence and the arousal reported after each topic were scored in comparison to this baseline. However, analyses were performed in order to substantiate this assumption.

In order to investigate linguistic alignment, the dialogues between the human and artificial agents were analyzed using Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004). Coh-Metrix is a tool that analyzes text based on nearly 1,000 different measures. Coh-Metrix goes beyond basic measures by using Latent Semantic Analysis (Landaur & Dumais, 1997) to look for the actual meaning of an utterance within a given context. Latent Semantic Analysis (LSA) uses a statistical algorithm to evaluate the semantic overlap of vectors within a text to that of indices from a large volume of corpora. The algorithm can evaluate the meaning of the words in context rather than simple keyword matching.

Linguistic alignment was calculated via Coh-Metrix which measures nearly 1,000 linguistic features. These indices have been reduced to five principle components which measure levels of cohesion denoting the ease to which readers create a connected mental representation of the text (Graesser, McNamara, & Kulikowich, 2011). For example, the representation may only be constructed by a basic understanding of the words used. On the other end of the comprehension continuum, the reader could fully understand the text

in relevance to the context in which it is presented. The similarity of the mental representation between the discourse of the student and the agent was used as a measure of linguistic alignment in the present research. More specifically, the principle components included narrativity (PC1), referential cohesion (PC2), syntactic simplicity (PC3), word concreteness (PC4), and causal cohesion (PC5) (Graesser et al., 2011). Narrativity refers to the amount of story-like features in the text. For example, more action words may be found in a more narrative type of speech. In the second PC of referential cohesion, there is a measure of the connectedness between the words and ideas of adjacent sentences in the text (in this case between the agent and the student). In the component of syntactic simplicity, there is continuum of many words with a complicated syntactic structure to few words with simple syntactic structures; this is indicative of analytical discourse and working memory load. The fourth component of word concreteness measures the extent to which there are abstract versus concrete words. Finally, the 5th PC of causal cohesion measures the goal-oriented connectives between the sentences and clauses within the text. This ability to articulate information in a cohesive manner represents a deeper level of understanding of the material (Graesser et al., 2011).

Measures of alignment were calculated for each component and topic individually by subtracting the overall score of the student from the overall score of the agent for that specific topic. For example, if the agent has a score for PC1 of 25 for Topic 1 and the student has a total score of 22, then the resulting value for alignment for Topic 1 is 2. In order to avoid negative values, because they are not indicative of further alignment, absolute values for each score were used. After computing the difference for each component, the scores were standardized before analysis.

Analyses

A series of linear mixed fixed-random effects models were performed in order to investigate the relationship between tutorial condition, mood contagion, linguistic alignment and learning gains on the item level ($N = 192$). For each of the following series of models, topic (Chapters 1 through 4), the test version (counter-balance order of pretest and posttest), and participant were entered as random factors to allow for greater generalization of results. First, mood condition was assessed in relation to prior-knowledge followed by learning. Then, valence and arousal were analyzed in relation to mood condition as well as learning. Finally, the linguistic features including the LIWC and Coh-Metrix scores were investigated in relation to the affective scores as well as learning.

Differences in prior-knowledge should have been distributed equally between conditions via random assignment. However, in order to assess this assumption before proceeding to analyses of learning gains, pretest scores were investigated. A model with condition as a fixed factor and participant, topic and test as random factors was used to assess pretest scores. It was not significantly different from a null model including the random factors of participant, topic and test and no fixed factors ($X^2(3, N = 192) = 1.02, p = .80$). These results suggest that prior-knowledge of the students was similarly distributed across the four mood conditions.

In order to determine whether differences in learning existed between the three conversational moods (positive, negative, neutral) and the control (neutral_{text only}), a mixed four-mood conditions (positive, negative, neutral, neutral_{text only}) and pretest fixed effects model with topic, test version, and participant held as random factors was used to

evaluate posttest scores. The model was not significantly different from the null model including pretest as a fixed factor and participant, topic, and test as random factors ($X^2(3, N = 192) = .50, p = .92$). This means that the students' learning was not altered by the mood conditions. Therefore, post hoc analyses were not performed. However, the reader may be interested to see the means and standard deviations of both pretest and posttest for each group which are found in Table 1. Though not significant, it does appear that the negative mood condition yielded the highest learning gains.

Table 1

Means of Pretest and Posttest Scores by Tutorial Condition

Column1	Pretest	Pretest	Posttest	Posttest
Condition	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
Text	0.55	0.22	0.65	0.23
Neutral	0.55	0.22	0.63	0.23
Negative	0.53	0.21	0.65	0.24
Positive	0.57	0.22	0.64	0.25

Next, a manipulation check was performed in order to ensure that the mood displayed by the pedagogical agents transferred to the participant. The measures used for mood contagion in this analyses were the 2 dimensional affect grid self-reports of valence and arousal completed by the participant. Before analyses were conducted, these values were standardized so that interactions between valence and arousal could be analyzed. Using a model including condition as a fixed effect and participant, topic and test as random factors, the manipulation appeared to have indeed induced reported valence. Specifically, the condition of the tutor (i.e., positive, negative, neutral, or control) had a

significant main effect on self-reported valence of the student ($F(3,189) = 5.63, p < .01$). The model was significantly different from the null model including the random factors of participant, topic and test ($\chi^2(3, N = 192) = 16.30, p < .001$) with a difference in variance accounted for of 8% ($R^2 = .076$). Therefore, the pedagogical agent's mood accounts for 8% of the variance in student reported affective valence. Post hoc analyses with a Tukey correction showed significant mood contagion for the negative and positive tutorial conditions, with a marginally significant difference from the neutral condition. Specifically, the negative condition showed an increase in negatively-valenced affect compared to the text (or control) condition ($z(1,192) = -3.987, p < .001$). The positive condition showed a significant increase in positively-valenced affect compared to the negative tutorial condition ($z(1,191) = 2.65, p < .05$), and the neutral tutorial condition showed a marginally significant increase in negative affect ($z(1,191) = 2.310, p < .1$). A table of the estimates of the pairwise comparisons can be seen in Table 2.

Table 2

Estimates of induced Valence compared to Text

Condition	Estimate	Std.Error
Neutral To Text	-0.39	0.17
Negative To Text	-0.67	0.17
Positive To Text	-0.23	0.17
Negative to Neutral	-0.29	0.17
Positive to Neutral	0.17	0.17
Positive to Negative	0.45	0.17

Comparable models as used to assess valence were used to analyze the effects of the pedagogical agents' mood on the interaction of valence and arousal as well as arousal individually. A model including mood condition (i.e., positive, negative, neutral and control) as a fixed factor and participant, topic, and test as random factors revealed that the mood condition did not induce an interaction of arousal and valence in the participant self-reports ($X^2(3, N = 192) = 1.68, p = .64$). Similarly, in assessing arousal, a full model including condition as a fixed factor and participant, topic and test as random factors did not significantly differ from a null model including the random factors only ($X^2(3, N = 192) = 5.72, p = .13$). Therefore, the condition of the teacher did not significantly induce arousal or the interaction of valence and arousal as indicated by the student's affective grids.

The third hypothesis investigated was the relationships between the students' valence, arousal, the interaction between the two, and learning gains. A series of additive models were compared to assess the whether any relationships were additive or interactive. The first model including valence and pretest as fixed factors and participant, topic and test as random factors was compared to the null model including pretest as a fixed factor and the three random factors. The model comparison revealed valence did not significantly impact learning as it is not different from the null model ($X^2(1, N = 293) = .02, p = .88$). Next, arousal was added creating a full model of arousal, valence and pretest as fixed factors with participant, topic, and test as random factors which was significantly different from the previous model ($X^2(1, N = 192) = 3.72, p = .05$). This suggests that arousal significantly contributes to learning. Finally, the interaction was tested with a model including arousal, valence, the interaction between the two and

pretest as fixed factors with participant, topic, and test as random factors. This model was not significantly different from the model including arousal, valence, and pretest as fixed factors and the three random factors ($\chi^2(1, N = 192) = .54, p = .46$). Therefore, the interaction of valence and arousal did not impact learning.

In the previous model comparisons, arousal significantly contributed to learning gains. These findings were revealed in a full model including valence, arousal, and pretest as fixed factors with participant, topic and test as random factors. Specifically, this model showed valence to contribute near 0% of the variance ($R^2 = .001$) whereas arousal accounted for 1.2% of the variance ($R^2 = .012$). There was a significant main effect for arousal ($F(1, 191) = 4.20, p < .05$) and pretest ($F(1, 191) = 4.25, p < .05$), but not valence ($F(1, 191) = .14, p = .71$). The relationship between arousal and learning is negative ($t(1, 191) = -2.02, p = .04$). Therefore, the lower the intensity of affect reported by the student, the higher the learning gains.

There is a potential criticism to these analyses of the impact of arousal and valence on learning. Specifically, the model comparisons did not show the statistical significance of each fixed factor compared to the null model. It is noteworthy to mention that a model including arousal and pretest as fixed factors and participant, topic, and test as random factors only maintained marginal significance when compared to a null model of pretest as a fixed factor and the three random factors ($\chi^2(1, N = 192) = 3.12, p = .08$). This finding is not surprising taking into account the small effect size of arousal on learning.

Another possible limitation is the calculation of the measures of arousal. These were performed under the assumption that each student would return to his or her original

baseline between each within-subjects condition. However, if one were to assume that this neutralization did not occur, then the change in valence and arousal can be calculated in comparison to the previous condition only. Under this comparison, both the effects of arousal on learning and tutorial condition on valence were found to be comparable to the results obtained under the original assumption.

Linguistic Measures

Coh-Matrix measures. The original hypothesis posited that the Coh-matrix measures of alignment and LIWC measures of mood contagion would moderate the effects of condition on learning. Though main effects of condition were not found, investigations continued into the relationship between linguistic alignment, student mood and learning. The following analyses concentrate on linguistic measures, so only the conversational conditions could be analyzed, thus removing the text condition and reducing the number of observations ($N = 144$).

In order to test the relationship between linguistic alignment and learning, a full model was constructed including the standardized alignment scores representing the 5 components and pretest as fixed factors, with participant, topic and test as random factors. The model is significant compared to a null model including pretest as a fixed factor and participant, topic, and test as random factors ($X^2(5, N = 144) = 15.12, p = .009$). Interestingly, 2 of the 5 scores for the principal components measuring alignment between student and agent correlated with learning. Specifically, the component of concreteness had a significant main effect ($F(1,143) = 10.15, p < .05, R^2 = .037$) on learning and accounted for 3.7% of the variance. Also, the standardized alignment score for component of causal cohesion showed a significant main effect in the full model

($F(1,143) = 3.89, p = .05$). However, the component only maintained marginal significance after the variance attributed to the concrete component was partialled out ($p = .06$) and contributed only 1.2% of the variance. Though standardized values were used in the mixed-effects models, the mean of the un-standardized alignment score showed the concreteness component to have a higher amount of linguistic alignment than any of the other components. A full list of means can be seen in Table 3. It is important to take into account that smaller values indicate higher alignment when interpreting the table.

Table 3

Means and Standard Deviations of Alignment Scores by Component

Score	Narrativity	Referential Cohesion	Syntactic Simplicity	Concreteness	Causal Cohesion
<i>MEAN</i>	37.96	56.5	22.09	9.23	32.49
<i>SD</i>	22.02	22.09	19.5	16.33	20.01

Counter-intuitively, the nature of the relationship between the alignment scores and learning suggests divergent alignment associated with learning ($t(1,143) = 3.09, p < .05$) ($t(1,143) = 1.98, p = .05$). Explanations for this finding are elaborated in the discussion section.

LIWC Measures. The LIWC Measures of positive and negative valence did not yield any significant findings in accounting for learning. Specifically, the full model containing pretest along with positive and negative words as fixed factors and participant, test, and topic as random factors did not significantly differ from the null model of pretest

as a fixed factor and participant and tests as random factors ($X^2(2, N = 144) = 1.62, p = .45$).

An additional investigation was conducted in order to see if the linguistic affective measures related to the valence and arousal reported in the affect grids. Similarly, the valence of the words used did not predict the valence reports on the affect grid when compared to a null model of participant, topic, and test version as random factors ($X^2(2, N = 144) = .98, p = .62$). Therefore, the affective linguistic measures do not correlate with the self-reports of valence.

A possible explanation is the small number of words that were analyzed per participant (overall mean word-count per unit of analysis is 78.4) which may not be enough for an accurate analysis. This may be especially the case in the tutorial context where such emotive words may not be frequently expressed by a student.

Discussion

The current study investigated multiple hypotheses about the nature of mood and learning. The hypothesis that the mood displayed by the artificial pedagogical agent would affect learning was unsupported in this study. The lack of significant findings cannot be attributed to the manipulation itself because the mood induction was successful. The tutorial conversations had a significant main effect on the valence of the student reported emotion. Specifically, when the artificial agent used multiple negatively-valenced words, the student reported a concurrent negative state as compared to baseline. Similarly, the positive words spoken by the pedagogical agent produced a more positively-valenced mood than the negatively-valenced tutorial condition. Interestingly,

the neutral mode of conversation seemed to trend towards producing a negative mood as well.

The mood of the teacher did not have a significant main effect on learning, but the student's self-reported affective state of arousal did have a small but significant main effect on learning. Specifically, the less arousal displayed by the student, the more learning was achieved. The most plausible explanation is the arousal theory suggesting that intensity of emotions rather than valence contribute to learning on certain tasks (Bradley et al., 2001; Lang, 1995). Previous research suggests that this may be the case when the task is didactic in nature (Martindale, 1981; Yerkes & Dodson, 1908). The four chapters covering the definition, importance, and example of basic topics of research methodology can certainly be described as didactic and therefore fall in line with this previous research.

However, perhaps there is a more intricate explanation offered by Pekrun's control-value theory (2006). Students who are performance-avoidance oriented will likely have more arousal corresponding to heightened levels of fear of failure. These students achieve less learning gains than mastery-oriented students who theoretically should have lower levels of arousal due to a decreased focus on fear of reprimand or negative feedback. Perhaps it is possible that lower arousal is indicative of mastery-oriented students who tend to achieve higher learning gains. However, the lack of the corresponding emotion does not substantiate this theory. Specifically, according to Pekrun's theory, mastery-oriented students should also show a higher level of positive affect leading to learning. However, in this study, this relationship between valence or an interaction of valence and arousal and learning was not discovered.

Linguistic alignment scores between the human student and the artificial pedagogical agents were also found to affect learning. The measure of word concreteness or use of meaningful words contributed significantly to this effect, whereas the measure for causal cohesion representing a deep-level understanding had a marginally significant main effect. The exact nature of the relationship between the alignment score of the concrete and causal cohesion components and learning gains is unclear. Specifically, one party, either student or artificial agent, is using more concrete words and possibly goal-oriented connectives than the other. From the current calculations, it cannot be determined whether it is the student or the agent producing such linguistic features. However, the mean alignment between student and agent was higher for the concrete score than any of the other linguistic features suggesting that a certain amount of alignment in the use of meaningful concretized words may be necessary for learning.

One possible conclusion from these results addresses the common ground hypothesis (Clark & Schaeffer, 1987). This hypothesis asserts that alignment is necessary only through the process of building common ground and then it dissipates. Therefore, perhaps the divergence in alignment between conversation partners is indicative of a level of previous convergent alignment necessary for achieving common ground. Specifically, in the tutorial context, it is clear that a necessary level of mutual understanding about the topic must be achieved in order for learning to occur. It is possible that common ground was achieved on the concrete level which may be all that is needed for the transfer of didactic knowledge. After the necessary level of common ground was attained, then the tutee could diverge from the linguistic features of the

artificial agents. Perhaps further investigations including a finer-grained analysis on the agent and turn level may reveal more insight into the exact nature of these findings.

Finally, the lack of significant findings using LIWC is not surprising due to the low number of words contributed per conversation. In future studies, longer interactions with tutorial agents may increase the chances of finding significant evidence of mood contagion within the language of the student. Elongated conversations may also make it possible to discover relationships between the mood of the animated pedagogical agent and a student along with additional information about the phenomenon of mood contagion and linguistic alignment.

The current study investigated the relationship between teacher mood, mood contagion, linguistic alignment and learning. Though the current study was unable to answer all of the proposed research questions, the overall finding can help guide future research. Empirical evidence which sheds light on the inner-workings of these constructs may help guide the development of an intelligent tutoring system which adapts to individual students' specific needs.

References

- Azevedo, R. (2007). Understanding the complex nature of self-regulated learning processes in learning with computer-based learning environments: An introduction. *Metacognition and Learning, 2*, 57-65.
- Baker, L. (1985). How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest-Pressley, G. E. MacKinnon, & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (pp. 155-205). Orlando, FL: Academic Press.
- Bavales, J.B., Black, A., Lemery, C.R., & Mullett, J. (1986). "I show how you feel": Motor mimicry as a communicative act. *Journal of Personality and Social Psychology, 50*, 322-329.
- Bernieri, F., Reznick, J.S., & Rosenthal, R. (1988). Synchrony, pseudo synchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions. *Journal of Personality and Social Psychology, 54*, 243-253.
- Bloom, B.S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, NY: McKay.
- Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1*, 276-298.
- Brown, P., & Levinson, S. (1978). Universals in language use: Politeness phenomena. In E.N. Goody, (Ed.), *Questions and politeness: Strategies in social interaction* (pp.56-289). New York, NY: Cambridge University Press.
- Burleson, W. (2006). *Affective learning companions: Strategies for empathetic agents with real-time multimodal affective sensing to foster meta-cognitive approaches to learning, motivation, and perseverance*. (Unpublished doctoral dissertation). MIT, Cambridge, MA.
- Byrne, D. (1971). *The attraction paradigm*. New York, NY: Academic Press.
- Cassell, J., & Tversky, D. (2005). The language of online intercultural community formation. *Journal of Computer-Mediated Communication, 10*, 29-59.
- Chartrand, T.L., & Bargh, J.A. (1999). The Chameleon Effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*, 893-910.

- Clark, H. H., & Schaefer, E. F. (1987). Collaborating on contributions to conversations. *Language and Cognitive Processes*, 2, 19-41.
- Clark, R.C., & Mayer, R.E. (2002). *E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. San Francisco; Jossey-Bass Pfeiffer.
- Craig, S.D., Graesser, A.C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29, 241-250.
- D'Mello, S., & Graesser, A.C. (in press). The half-life of cognitive-affective states during learning. *Cognition & Emotion*.
- 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, 187-119.
- D'Mello, S.K., Picard, R., & Graesser, A.C. (2007). Toward an affect-sensitive AutoTutor. *IEEE Intelligent Systems*, 22, 53-61.
- Duclos, S.E., Laird, J. D., Schneider, E., Sexter, M., Stern, L., & Van Lighten, O. (1989). Emotion-specific effects of facial expressions and postures on emotional experience. *Journal of Personality and Social Psychology*, 57, 100-108.
- Duncker, K. (1945). On problem-solving. *Psychological Monographs*, 58, (Whole No.5).
- Eisenhower, R.W., Frank, T.D., Kay, B.A., & Carello, C. (2010). A dynamical systems approach to emotional experience: Relating affective and event valence. Paper presented at 20th Annual New England Sequencing and Timing Conference: New Haven, CT.
- Franzel, A.C., Goetz, T., Ludtke, O., Pekrun, R., & Sutton, R.E. (2009). Emotion transmission in the classroom: Exploring the relationship between teacher and student enjoyment. *Journal of Educational Psychology*, 101, 705-716.
- Frederickson, B.L. (1998). What good are positive emotions? *Review of General Psychology*, 2, 300-319.
- Gholson, B., & Craig, S. D. (2006). Promoting constructive activities that support learning during computer-based instruction. *Educational Psychology Review*, 18, 119-139.
- Giles, H., Fortman, J., Dailey, R., Barker, V., Hajek, C., Anderson, M.C., & Rule, N., (2005). Communication accommodation: Law enforcement and the public. *Center*

on *Police Practices and Community (COPPAC)*. Retrieved October 27, 2006 from http://repositories.cdlib.org/isber/coppac/articles_chapters/.

- Glasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science, 13*, 33-39.
- Glenberg, A.M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 702-718.
- 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., Conley, M., & Olney, A. (2012). Intelligent tutoring systems. In K.R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Applications to learning and teaching* (pp. 451-473). Washington, DC: American Psychological Association.
- Graesser, A. C., Jeon, M., & Dufty, D. (2008). Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes, 45*, 298-322.
- 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., McNamara, D.S., & Kulikowich, J. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*, 223-234.
- Graesser, A.C., McNamara, D.S., Louwerse, M.M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers, 36*, 193-202.
- Halpern, D.F., Millis, K., Graesser, A.C., Butler, H., Forsyth, C., & Cai, Z. (in press). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity*.
- Hatfield, E., Cacioppo, J. L., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Sciences, 2*, 96-99.
- Hatfield, E., & Rapson, R. L. (2004). Emotional contagion: Religious and ethnic hatreds and global terrorism. In L. Z. Tiedens & C. W. Leach (Eds.), *The social life of emotions* (pp129-143). Cambridge, UK: Cambridge University Press.
- Howes, M.J., Hokanson, J.E., & Lowenstein, D.A. (1985). Induction of depressive affect after prolonged exposure to a mildly depressed individual. *Journal of Personality and Social Psychology, 49*, 1110-1113.

- Hsee, C.K., Hatfield, E., & Carlson, J.G.(1990). The effect of power on susceptibility to emotional contagion. *Cognition and Emotion*, 4, 327-340.
- Isen, A.M. (2008). Positive affect and decision processes: Some recent theoretical developments with practical implications. In C.Haugdevedt, P.M. Herr, & F.R. Kardes (Eds), *Handbook of consumer psychology* (pp. 273-296). New York: Psychology Press.
- Isen, A.M., Daubman, K.A., & Nowicki, G.P.(1987). Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology*, 52, 1122-1131.
- Jackson, G.T., Graesser, A.C., & McNamara, D.S. (2009). What students expect may have more impact than what they know or feel. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A.C. Graesser (Eds.), *Artificial intelligence in education; Building learning systems that care; From knowledge representation to affective modeling* (pp. 73-80). Amsterdam, The Netherlands: IOS Press.
- Jung, C. G. (1968). *Lecture five. Analytical psychology: Its theory and practice*. New York, NY: Random House.
- Jurafsky,D., & Martin, J.(2008). *Speech and language processing*. Englewood, NJ: Prentice Hall.
- Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Teachers' occupational well-being and the quality of instruction: The important role of self-regulatory patterns. *Journal of Educational Psychology*, 100, 702–715.
- Lakin, J.L., Jefferis, V.E., Cheng, C.M., & Chartrand, T.L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, 27, 145-162.
- Landauer, T.K., & Dumais, S.T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.
- Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologist*, 50, 372–385.
- Linnenbrink, E. A., & Pintrich, P. R. (2002). Achievement goal theory and affect: An asymmetrical bidirectional model. *Educational Psychologist*, 37, 69–78.
- Martindale, C. (1981). *Cognition and Consciousness*. Homewood, IL: Dorsey Press.
- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers*, 36, 222-233.

- Mednick, M.T., Mednick, S.A., & Mednick, E. (1964). Incubation of creative performance and specific associative priming. *Journal of Abnormal and Social Psychology, 69*, 84-88.
- Mehi, M., & Pennebaker, J. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversation. *Journal of Personality and Social Psychology, 84*, 857-870.
- Millis, K, Forsyth, C., Butler, H., Wallace, P., Graesser, A., & Halpern, D. (2011) Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou & J. Lakhmi (Eds.) *Serious games and edutainment applications* (pp.169-196). London, UK: Springer-Verlag.
- Moreno, R., & Mayer, R.E. (2002). Verbal redundancy in multimedia learning: when reading helps listening. *Journal of Educational Psychology, 94*, 156-163.
- Niederhoffer, K.G., & Pennebaker, J.W. (2002). Linguistic style-matching in social interaction. *Journal of Language and Social Psychology, 21*, 337-360.
- Neumann, R., & Strack, F. (2000). Mood contagion: The automatic transfer of moods between persons. *Journal of Personality and Social Psychology, 79*, 211–223.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315-341.
- Pekrun, R., Maier, M.A., & Elliot, A.J. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101*, 115-135.
- Pennebaker, J.W., Francis M.E. & Booth R.J. (2001). *Linguistic inquiry and word count (LIWC):LIWC2001*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Reeves, B., & Nass, C.(1996). *The media equation*. New York, NY: Cambridge University Press.
- Reik, T. (1948) *Listening with the third ear: The inner experience of a psychoanalyst*. New York, NY: Farrar, Straus, and Giroux.
- Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172.
- Russell, J.A., Weiss, A., & Mendelsohn, G.A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*,493-502.

- Schimmack, U. (2005). Attentional interference effects of emotional pictures: Threat, negativity, or arousal?. *Emotion*, 5, 55-66.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, 45, 513–523.
- Scissors, L.E., Gill, A.J., & Gergle, D. (2008). Linguistic mimicry and trust in text-based CMC. In B. Begole & D.W.McDonald (Eds.) (2008). *Proceedings of Computer Supported Cooperative Work* (pp. 277-280). New York, NY: ACM Press.
- Sullins, J., Craig, S., & Graesser, A.C. (2009). Tough love: The influence of an agent's negative affect on students' learning. In V. Dimitrova, R. Mizoguchi, B. Du Boulay, & A.C. Graesser (Eds.), *Artificial intelligence in education: Building learning systems that care: From knowledge representation to affective modeling* (pp. 677-679). Amsterdam: IOS Press.
- Sy, T., Côté, S., & Saavedra, R. (2005). The contagious leader: Impact of the leader's mood on the mood of group members, group affective tone, and group processes. *Journal of Applied Psychology*, 90, 295–305.
- Tickle-Degnen, L., & Rosenthal, R. (1987). Group rapport and nonverbal behavior. *Review of Personality and Social Psychology*, 9, 193-220.
- 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.
- Yerkes R.M., & Dodson J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459–482.
- Zeidner, M. (2007). Test anxiety in educational contexts: Concepts, findings, and future directions. In P. A.Schutz & R. Pekrun (Eds.), *Emotion in education* (pp.13–36). Amsterdam: Elsevier.

APPENDIX A

Example Script

This is an example script highlighting the scaffolding methods used by the artificial agents during interaction with a real human.

QUESTION: Glass: Okay, I know operational definitions are important in research, but I'm not sure why.

HINT: Glass: Let's just go over the importance of these things one more time. Why do we need to have operational definitions?

Human: so that they can be reliably measured and used by other researchers

NEUTRAL FEEDBACK: Glass: OK.

PUMP: Dr. Quinn: Um, can you add to that?

Human: it is important to create operational definitions because they provide a basis for measuring variables and so that they can be reliably measured and used by other researchers in other studies.

Glass: That is not clear to me.

HINT: Glass: I thought operational definitions had something to do with telling others about your study. What could that be?

Human: how to measure a variable

Glass: I still don't understand.

PROMPT: Glass: Wow, it seems like operational definitions have something to do with being able to reliably do something . What is that something?

Human: measure a variable

NEUTRAL FEEDBACK: Dr. Quinn: Okay.

CORRECT PROMPT COMPLETION: Dr. Quinn: reliably recognize variables

PROMPT: Glass: I'm remembering something about how operational definitions make it so that a variable means the same thing to all researchers. That means that when other researchers use the same variable, it is what?

Human: reliable

POSITIVE FEEDBACK: Dr. Quinn: Great!

Dr. Quinn: it is reliable

Glass: Yeah, I understand.

SUMMARY: Glass: Operational definitions are important to consider because they allow a particular variable to be reliably recognized, measured, and understood by all researchers.

APPENDIX B

Example Conditions

The positive, negative, and neutral conditions are shown in interactions with *Operation ARIES!*

The altered words are in bold.

Positive:

Dr. Quinn: I think it might be a **fantastic** idea for us to go over this **exciting** topic a little more so that we all understand.

Dr. Quinn: Here is an **engaging** question. Why don't you explain what you think an operational definition is?

Human: its a fact that can be proven true

Glass: I do not understand.

Dr. Quinn: Consider this.

Dr. Quinn: When considering the meaning of an operational definition, think about defining important things in an **amazing** study.

Human: it defines things in a study

Glass: I do not understand.

Dr. Quinn: Alright.

Dr. Quinn: Try this.

Dr. Quinn: I am **positive** that operational definitions are related to variables. For variables, operational definitions help researchers do what?

Human: find variables

Dr. Quinn: Okay.

Dr. Quinn: describe them

Dr. Quinn: See if you know the answer to this.

Dr. Quinn: By the way, I am feeling **fabulous** about this session. An operational definition is a way to identify, describe and measure variables in what kind of way?

Human: in experiments

Dr. Quinn: Alright.

Dr. Quinn: in a precise and agreed upon way

Dr. Quinn: An operational definition is a **great** way to identify, describe and measure variables in a precise and agreed upon way. For example, in a study looking at whether or not listening to classical music makes babies smarter, the researcher needs to define what is meant by "smart". One way to operationally define "smart" is by counting the number of words babies can say at 18 months of age.

Negative :

Dr. Quinn: **Unfortunately**,I think we might need to go over this **difficult** topic a little more so that we all understand.

Dr. Quinn: I know that discussing this may be in **vain**, but why don't you explain what you think an operational definition is?

Human: An operational definition measures variables in a particular study

Human: something about a study

Dr. Quinn: When seriously considering the **hellish** meaning of an operational definition, think about defining things in a **boring** study.

Human: not sure

Dr. Quinn: Operational definitions are related to **boring** variables. For variables, operational definitions help researchers do what?

Human: define things

Dr. Quinn: Alright.

Dr. Quinn: An operational definition is a **frustrating** way to identify, describe and measure variables in a precise and agreed upon way. For example, in a study looking at whether or not listening to classical music makes babies smarter, the researcher needs to define what is meant by "smart". One way to operationally define "smart" is by counting the number of words babies can say at 18 months of age.

Neutral:

Dr. Quinn: I think it might be a good idea for us to go over this topic a little more so that we all understand.

Dr. Quinn: Why don't you explain what you think an operational definition is?

Human: Operational definitions provide a way of defining and controlling variables in a particular study

Dr. Quinn: Um, anything else?

Human: they give another experimenter a way of replicating the experiment

Glass: I still don't understand.

Dr. Quinn: Okay.

Dr. Quinn: Let me help you a little.

Dr. Quinn: When trying to understand what operational definitions are, think about what is being defined.

Human: a variable

Dr. Quinn: Nice job!

Glass: Yeah, I understand.

Glass: An operational definition is a way to identify, describe and measure variables in a precise and agreed upon way. For example, in a study looking at whether or not listening to classical music makes babies smarter, the researcher needs to define what is meant by "smart". One way to operationally define "smart" is by counting the number of words babies can say at 18 months of age.

APPENDIX C

Counter-balanced Conditions

Assessment Version A Version B
Conditions

Positive, Negative, Neutral, Text	Positive, Negative, Text, Neutral	Positive, Text, Negative, Neutral	Positive, Text, Neutral, Negative	Positive, Neutral, Negative, Text	Positive, Neutral, Text, Negative
Negative, Positive, Neutral, Text	Negative, Positive, Text, Neutral	Negative, Text, Positive, Neutral	Negative, Text, Neutral, Positive	Negative, Neutral, Text, Positive	Negative, Neutral, Positive, Text
Neutral, Positive, Negative, Text	Neutral, Positive, Text, Negative	Neutral, Text, Positive, Negative	Neutral, Text, Negative, Positive	Neutral, Negative, Text, Positive	Neutral, Negative, Positive, Text
Text, Positive, Neutral, Negative	Text, Positive, Negative, Neutral	Text, Negative, Positive, Neutral	Text, Negative, Neutral, Positive	Text, Neutral, Positive, Negative	Text, Neutral, Negative, Positive

Assessment Version B Version A

Conditions

Positive, Negative, Neutral, Text	Positive, Negative, Text, Neutral	Positive, Text, Negative, Neutral	Positive, Text, Neutral, Negative	Positive, Neutral, Negative, Text	Positive, Neutral, Text, Negative
Negative, Positive, Neutral, Text	Negative, Positive, Text, Neutral	Negative, Text, Positive, Neutral	Negative, Text, Neutral, Positive	Negative, Neutral, Text, Positive	Negative, Neutral, Positive, Text
Neutral, Positive, Negative, Text	Neutral, Positive, Text, Negative	Neutral, Text, Positive, Negative	Neutral, Text, Negative, Positive	Neutral, Negative, Text, Positive	Neutral, Negative, Positive, Text
Text, Positive, Neutral, Negative	Text, Positive, Negative, Neutral	Text, Negative, Positive, Neutral	Text, Negative, Neutral, Positive	Text, Neutral, Positive, Negative	Text, Neutral, Negative, Positive

APPENDIX D: Example Questions for Pre-Test and Post-Test

The following questions are in descending order from shallow to deep-level questions.

Recognition

1. What does an operational definition provide?
(This is a homo-morphic question)
 - a. **A way to identify variables in a precise and agreed upon way.**
 - b. A way of using principles to guide research
 - c. A way to alter the results of a study
 - d. The use of variables in a study

2. What is the function of an operational definition?
 - a. To define medical equipment so that doctors can use it correctly
 - b. **To define variables so that other researchers can understand an experiment.**
 - c. It is to define variables so that other researchers can control the study.
 - d. It is to define objects so that they can be identified.

3. Which statement provides the best description of an operational definition?
 - a. A way of describing variables in a theoretically motivated fashion.
 - b. A way of describing the relationship between variables.
 - c. The use of simple arithmetic principles as a treatment effect.
 - d. **A way of recognizing and describing a variable in a particular study.**

4. Which statement best reflects the reason why it is important to have operational definitions?
 - a. Operational definitions enable researchers to distinguish between correlational and experimental variables.
 - b. **Operational definitions provided a basis for describing, manipulating, or measuring variables that can be easily understood and replicated by other researchers.**
 - c. Operational definitions provide a basis for creating descriptions of the variables in a study.
 - d. Operational definitions provide a basis for conducting experiments in psychology, but actually have little bearing on other sciences.

5. Which is the best example of an operational definition?
 - a. **A healthy heart is defined in terms of the amount of plaque in the arteries that supply the heart with blood.**
 - b. A healthy heart is defined in terms of a precise medical description of the physiology of the heart.
 - c. A healthy heart is defined in terms of whether someone lives a healthy lifestyle
 - d. A healthy heart is defined in terms of how many miles a day someone walks

6. A researcher decides to test the hypothesis that women are stronger than men. The researcher has open access to multiple locations on a college campus. Please write out a possible operational definition that could be used to define “strong”.
- how much weight a person can lift**
 - how well they do on a test
 - the women can be tested by muscle mass and the men can be tested by how much weight they can lift
 - both a and c

7. Please read the following hypothesis and answer the question:

(This is an isomorphic question)

A researcher decides to test the hypothesis that people who read a lot of books are smarter than people who do not read a lot of books.

Please read the following items and decide which would be the best possible operational definition for “people who read a lot of books”, could be:

- Counting the number of people who wear glasses because that probably means that they read a lot.
 - Deciding that all females read more books and therefore counting the number of females in the study.
 - With the subjects’ permission, check the library records for the number of books checked out over the last month**
 - Try to sell the subjects a specific book and the ones who buy it obviously read more.
8. A medical researcher decides to test to see if a pill will lower sugar levels in patients with diabetes. He only recruits patients to be subjects if they have diabetes. He also has one group that is given the new pill and another group which is given a placebo, which is defined as a pill with no medicine in it. He knows that there are normal fluctuations in sugar throughout the day, so he defines a lower sugar level as a persistently lower level of sugar by a significant amount which is usually $1/8$ of the patient’s normal level over 30 days. All patients are given the pill for 60 days and tested on a daily basis. Please identify where the researcher operationally defined a variable in the above scenario. **(Recall Question)**

THE UNIVERSITY OF MEMPHIS Institutional Review Board

To: Carolyn M. Forsyth
Psychology

From: Chair or Designee, Institutional Review Board
For the Protection of Human Subjects
irb@memphis.edu

Subject: ARIES and Emotions (060311-547)

Approval Date: June 22, 2011

This is to notify you that the Institutional Review Board has designated the above referenced protocol as exempt from the full federal regulations. This project was reviewed in accordance with all applicable statutes and regulations as well as ethical principles.

When the project is finished or terminated, please submit a Human Subjects Research Completion Form (COMP) to the Board via e-mail at irbforms@memphis.edu. This form can be obtained on our website at <http://www.memphis.edu/irb/forms.php>.

Approval for this protocol does not expire. However, any change to the protocol must be reviewed and approved by the board prior to implementing the change.

Chair or Designee, Institutional Review Board
The University of Memphis

Cc: Dr. Arthur Graesser