Journal of Educational Multimedia and Hypermedia (2018) **27**(2), 245-265

When Knowledge Isn't Power: The Influence of Prior Knowledge on Question Generation Training

JEREMIAH SULLINS
Harding University, USA
jsullins@harding.edu

SAMUEL ACUFF
University of Memphis, USA
sfacuff@memphis.edu

DANIEL NEELY
Auburn University, USA
din0001@tigermail.auburn.edu

XIANGEN HU
University of Memphis, USA
xiangenhu@gmail.com

Is it possible to teach a learner to become a better question asker in as little as 25 minutes? Questions are believed to play a crucial role in a variety of cognitive faculties, including comprehension and reasoning. Available research suggests that learning how to ask good questions should be taught at an early age but all ages benefit from question generation training. Sadly, consistent with the research coming out of self-regulation, it is well documented that the ideal scenario of a curious question asker does not match reality. Students are unspectacular at monitoring their own knowledge deficits and their question generation is both infrequent and unsophisticated. Given that many teachers and school districts do not have the resources to provide individualized question training to students, the current study sought to explore the benefits of using animated pedagogical agents to teach question asking skills in a relatively short amount of time (approximately 25 minutes). Results revealed a significant difference in the quality of questions generated on the posttest as a function of condition. Additionally, results revealed a significant interaction between prior knowledge and question training on the posttest scores.

Keywords: computer-assisted instruction, question asking, tutoring, human-computer interaction, vicarious learning

Question generation has received a great deal of attention in recent years from researchers in the fields of computer science (Heilman & Smith, 2010), psychology (Graesser, Ozuru, & Sullins, 2009; Rus & Graesser, 2009; Sullins & McNamara 2009) and education. However, Rosenshine et al. (1996) have stated that at the present time, developing procedural prompts (i.e., prompts to stimulate generation of different question types) appears to be an art. (p. 198). Although a fair amount of research does exist in the realm of question asking and procedural prompts, very little empirical work has been published to evaluate the benefits of different procedural prompts.

Question generation is believed to play a crucial role in a variety of cognitive faculties, including comprehension (Collins, Brown, & Larkin, 1980; Graesser, Singer, & Trabasso, 1994) and reasoning (Graesser, Baggett, & Williams, 1996; Sternberg, 1987). Asking good questions has been shown to lead to improved memory and comprehension of material in school children and adult populations (Rosenshine, Meister, & Chapman, 1996). Available research suggests that learning how to ask good questions should be taught at an early age, but all ages benefit from question generation training (Wisher & Graesser, 2007). Generally speaking, the overarching view has been that self-generation of questions produces deeper cognitive engagement such as more active reading, greater focus of attention on and elaborating of content, and greater self-awareness of the degree of comprehension. Bugg and McDaniel (2012) posit that any technique that stimulates more active processing (i.e., question generation) of text significantly improves metacomprehension accuracy. This is good news because for some time it has been known that students experience illusions of knowing and learning (Eva et al., 2004; Jee et al., 2006). This is consistent with other research that has found that students have insufficient mental resources to learn and monitor their own learning (Moos & Azevedo, 2008). Furthermore, it is well documented that the ideal scenario of a curious question asker does not match reality. Students are unspectacular at monitoring their own knowledge deficits and their question generation is both infrequent and unsophisticated (Baker, 1979; Dillon, 1988; Graesser & Person, 1994; Van der Meij, 1988).

Graesser and Person (1994) reported that an individual student asks approximately 1 question in 7 hours of class time (around 1 question per day). Most of these questions are not good questions, so the quality is also disappointing. More recent research has found similar results. A study examining approximately 10,000 assessment items from 50 instructors in the U.S. discovered that over 90% of questions were at the lowest two levels of Bloom's taxonomy (Momsen, Long, Wyse, & Ebert-May, 2010).

It is well documented that students and adults have trouble generating questions (Dillon, 1988; Graesser & Person, 1994; Wisher & Graesser, 2007). Of the questions that are generated, the majority are shallow questions rather than questions that require deep reasoning. A deep reasoning question is one which integrates content and fosters understanding of the components and mechanisms being covered (Craig, Vanlehn, & Chi, 2009). Deep reasoning questions are questions that typically invite lengthier answers (usually around a paragraph in length) and often start with words such as *why, how,* or *what-if* (Graesser, Ozuru, & Sullins, 2009). These questions are aligned with the higher levels of Bloom's taxonomy (1956) and the long-answer question categories in the question taxonomy proposed by Graesser and Person (1994). In order to illustrate the difference between shallow reasoning questions and deep reasoning questions, consider an example of each. An example of a shallow reasoning question, according to Graesser and Person (1994), would be "Does the CPU use RAM when running an application?" The reason for categorizing this type of question as "shallow" is because it does not require substantial thought on the student's part; indeed, the student could answer it by simply guessing yes or no. In contrast, a deep reasoning question would be "How does the CPU use RAM when running an application?" The reason for categorizing this question as "deep" is because the student must use the knowledge known about computers to articulate the causal mechanisms that relate two components in the operating system. They not only need to generate a nontrivial amount of content, but must also be able to reason about complex causal mechanisms.

LEARNER GENERATED QUESTIONS

Li et al. (2014) investigated question asking during collaborative problem solving in an online game environment, Land Science. The study included 100 middle and high school students that participated in seven Land Science games. Researchers examined the transcripts of the interaction between participants during these games and manually identified 1,936 questions from student chats, and coded them as deep or shallow based on the Grasser and Person (1994) taxonomy. It was discovered that players did ask more questions as task difficulty and task unfamiliarity increased. However, consistent with previous research, results revealed that shallow questions were significantly more frequent than deep or intermediate questions. In other words, these findings confirmed that question asking during a collaborative educational game are similar to a typical classroom environment.

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

Based on these findings it can be argued that students need to be receiving question asking training outside of the classroom. In order to examine the degree of question training that children are receiving outside of the classroom, Sullins, Howard, and Goza (2014) investigated the quantity and quality of questions asked during popular children's educational shows. The shows that were selected were aired with regularity on popular television channels marketed towards children (e.g., Sprout and Disney Junior).

Researchers gathered available transcripts of each show averaging 230-290 minutes per show. Once all transcripts had been collected and checked for time equivalency, three independent researchers examined all transcripts and marked where a question had been asked during the episode. Researchers collected approximately 4,200 questions from 32 hours of shows. All questions were then analyzed as "deep" or "shallow" based on the Graesser and Person (1994) question taxonomy. Results revealed a healthy distribution of both deep-level reasoning and shallow-level reasoning questions. However, not all shows are created equal when it comes to the availability of question training techniques. More specifically, four of the seven shows had a distribution of deep questions at 28% or lower. In other words, during four of the seven shows, the majority of questions that are being asked during these episodes are shallow questions. These findings imply that not only are students not being exposed to deep questions within the classroom, but children are not receiving the question training/exposure to good deep reasoning questions outside of the classroom.

It is evident from the previously mentioned research that questions are a crucial component to the learning process. In fact, a deeper understanding can be achieved by having students pose, and answer, questions (Chin & Brown, 2002; Rosenshine et al., 2006; Draper, 2009). However, it is also evident from the previously mentioned research that children and adults alike have difficulty generating deep reasoning questions that lead to a deeper understanding of information.

EFFECTS OF PRIOR KNOWLEDGE OF QUESTION GENERATION

There is little doubt that the level of prior knowledge has an impact on various aspects of a student's learning and cognition (e.g., Song, Kalet, & Plass, 2016; Bringula, Basa, & Dela Cruz, 2016; Fyfe & Rittle-Johnson, 2016). However, much less is known regarding the impact of prior knowledge when it comes to various procedural prompts such as student question generation. One study conducted by Hardy et al., (2014) examine the relationship between student-generated content (i.e., multiple choice questions) and achievement. Results revealed benefits for students of the highest and lowest ability. However, it was the students with lower/intermediate ability who may have benefitted most. The authors do mention that their analysis did not take into account the *quality* of student questions and that a fruitful avenue for future work might be to investigate how quality of questions may help promote deep learning.

VICARIOUS LEARNING

Given that teachers and school districts do not have the resources available to provide individual question training, one area of research that gives rise to optimism is the area of vicarious learning. According to constructivist epistemology, learners actively create meaning and knowledge by interacting with people and other objects. Rather than simply delivering information, learning environments should stimulate the learner to actively construct knowledge. This traditional view of constructivism has focused on keeping the learner physically active, usually by interactivity. However, Mayer has made the claim that the learner does not have to be physically active in order for constructivist learning to occur. In other words, under this view, the learner must only be actively processing (i.e., cognitively active) during knowledge acquisition (Harvard, 2014).

In this situation, the physically passive learner would engage in a form of vicarious learning. For our purposes, vicarious learning is defined as learning in multimedia environments under conditions in which the user is passive, in that they do not physically interact in any way with the source of the information. Historically, the term vicarious learning was frequently used synonymously with observational learning, social learning, or modeling (e.g., Bandura, 1962; Rosenthal & Zimmerman, 1978). According to this perspective, by simply observing activities carried out by others, learners can master those activities without overt practice or direct incentives (Rosenthal & Zimmerman, 1978).

Current trends in educational technology such as computer-based courses (e.g., Mayer, 2009; Sitzmann, 2011) and distance learning (Bourdeau & Bates, 1997; Moore, Dickson-Deane, Galyen, 2011) have created situations in which learners are more and more likely to find themselves trying to gain knowledge as observers (Cox, McKendree, Tobin, Lee, & Mayers, 1999) rather than active participants. Because of these technologies, further empirical understanding of the conditions that promote learning among relatively isolated observers is required.

Available research has compared student learning gains in the context of vicarious learning environments versus interactive learning environments. For example, Craig, Sullins, Witherspoon and Gholson (2006) conducted two experiments in order to compare student learning gains between interacting and observing. In Experiment 1 students were randomly assigned to one of five different conditions (one interactive and four vicarious). Students in the interactive condition interacted with an intelligent tutoring system called AutoTutor. The learners in this condition used a dialogue box and a keyboard to respond to AutoTutor's spoken questions, assertions, hints, prompts, pumps, back-channel feedback and gestures. The video and audio of each interactive session was recorded. In one vicarious learning condition, each recorded interactive session was presented to a yoked participant who simply watched and listened to it. This condition was known as the yoked-vicarious. A second vicarious condition was simply a monologue that contained the same information as in the interactive and yoked-vicarious condition. The information was presented using the same voice engine and agent. This condition was known as monologue-vicarious. In a third vicarious condition, half of the "main points" were preceded by a deep level reasoning question. In the context of AutoTutor, these "main points" are known as ideal answers and expectations. In this vicarious condition, only the ideal answers were preceded by deep level reasoning questions. This condition was known as half-question vicarious. A fourth and final vicarious condition included deep level reasoning questions that preceded every ideal answer and expectation in the monologue. This condition was known as full-questions vicarious. Results revealed that learners in the full-questions vicarious condition significantly outperformed learners in each of the other four conditions.

In Experiment 2, participants were randomly assigned to one of four different conditions: interactive, yoked-vicarious, full-questions vicarious presented as a monologue, and full-questions vicarious presented as a dialogue. As in Experiment 1, participants in the interactive condition directly interacted with AutoTutor on 12 topics concerned with computer literacy. The video and audio of the interactive condition was recorded and showed to participants in the yoked-vicarious condition. The full-questions vicarious presented as a monologue condition contained deep level reasoning questions before every ideal answer and expectation. In this condition, the same agent and voice engine used in the interactive and yoked vicarious condition spoke the question and content. In the full-questions vicarious presented as a dialogue condition every ideal answer and expectation were preceded by a deep level reasoning question. However, the deep level reasoning question was asked by a separate distinct voice. Only the agent from the previous three conditions was present on the screen. Results revealed that both vicarious deep level reasoning question conditions significantly outperformed both the yoked-vicarious and interactive conditions.

The available research suggests that asking good questions is an efficient learning strategy that can lead to longer retention and a deeper understanding of information. Research also shows that under the right conditions, learners do not need to be physically engaged with the material to be learned.

COGNITIVE LOAD THEORY

For decades much research has revolved around Sweller's cognitive load theory. This theory states that humans have a limited working capacity and any instruction needs to take our limited working memory into consideration. Sweller states that instructional material contains three specific types of cognitive load: 1) extraneous cognitive load, which is generated by the manner in which information is presented to the learner (i.e., the design), 2) intrinsic cognitive load, which is the idea that all instruction has an inherent difficulty associated with it and 3) germane cognitive load which is the load dedicated to the processing, construction and automation of schemas.

Taking the idea of cognitive load theory, Mayer has postulated similar ideas around the development of multimedia instruction. Mayer has developed his theory of multimedia learning around the belief that people have a limited working memory capacity, that we process information using a dual channel system (auditory and visual) and that we must be actively processing the information that is presented. Similar to the three types of cognitive load presented by Sweller, Mayer postulates that three types of cognitive load must be managed when presenting multimedia instruction: 1) extraneous processing refers to cognitive processing that does not support the objective of the lesson. This is caused by poor instructional design. Examples of extraneous processing would be daydreaming by the learner or a professor is doing something unusual during the presentation of information and the learner is paying attention to the unusual instead of the information. 2) essential processing which is the basic cognitive processing required to mentally represent the presented material. This is caused by the inherent complexity of the material. 3) generative processing is the deep cognitive processing required to make sense of the presented material (Harvard, 2014). This is caused by the learner's motivation to make an effort to learn (e.g., relating information to prior knowledge). Based on these principles of cognitive load the goal is to reduce extraneous processing, manage essential processing and foster generative processing. In other words, cognitive capacity > = extraneous + essential + generative.

CURRENT STUDY

In order to address the gap in the existing literature, the present study sought to answer the question "Is it possible to teach a learner to become a better question asker in as little as 25 minutes?"

Predictions

According to *vicarious learning theory* (Sullins, Craig, & Graesser, 2010) learners who receive information vicariously should outperform learners who receive no training. A vicarious learning environment is one in which learners are not the addressee of the material and/or they do not have control over the material they are expected to master. Previous research has found vicarious learning environments to be an effective source of information delivery that significantly increases students' learning when compared to various controls (Gholson & Craig, 2006; Muller & Sharma, 2012). Based on vicarious learning theory, it would be expected that the learners in the question training condition would significantly outperform the learners in the control condition (question training > control).

There is reason to believe that a learner's level of prior knowledge would impact their pretest to posttest performance. *Cognitive load theory* (Sweller, 1988) states that we have a limited working memory capacity. If multimedia learning environments contain too much information, during schema development, low knowledge learners may experience a bottleneck of information which could prohibit any learning from taking place. The high knowledge learners could activate preexisting schemas that would offset the poor design and learn regardless of the interface limitations (high knowledge question training > low knowledge questions training = control).

However, scaffolding may help overcome limited working memory capacity in order to achieve schema development in low knowledge learners. The scaffolding that is necessary for the low knowledge learners may be detrimental to learners with high prior knowledge due to the redundancy of the information. According to the *expertise reversal effect* (Kalyuga, Ayers, Chandler, & Sweller, 2003), question training would be most beneficial to the low knowledge learners and may in fact hurt the performance of high knowledge learners (low knowledge > high knowledge = control).

Vicarious Learning Environment

The AutoTutor LITE system is based on AutoTutor (Graesser, Chipman, Olney, & Haynes, 2005). One challenge of the original AutoTutor system was its scalability due to its dependence on language analyzers. The version of AutoTutor LITE used in this study is a minimalist implementation of AutoTutor. It only includes an AutoTutor style interface and interaction with a lightweight language analyzer. This provides the learner with a streamlined tutorial interaction that relies on tutor hints and feedback for tutoring on a coherent brief chunk of information called a Shareable Knowledge Object or SKO (Hu et al., 2014).

Similar to AutoTutor, Autotutor LITE interacts with students using natural language and is most effective when the learning objectives are qualitative/conceptual. AutoTutor LITE requires users to construct an answer to the question. A typical system interaction starts with a general seed question. The system evaluates the student's answer and asks follow up questions, which it selects based on the student model. AutoTutor LITE provides feedback and selects the next questions based on the four indices of Learners Characteristics Curves. The current implementation of AutoTutor LITE uses extended weighted keyword matching and latent semantic analysis. See Figure 1 for an example screen capture of AutoTutor LITE.

One of the challenges of an ITS is creating a student model (Graesser, Person, Harter & TRG, 2001) that adequately assesses a student's knowledge. For example, an experienced human tutor can estimate how much a student knows or does not know by evaluating a student's answers to key questions. The human tutor can provide feedback to help a student actively construct responses that are relevant to the questions asked. AutoTutor LITE used LCC as a student model to offer appropriate feedback. For the purpose of the current study, AutoTutor Lite was used as an information delivery system.

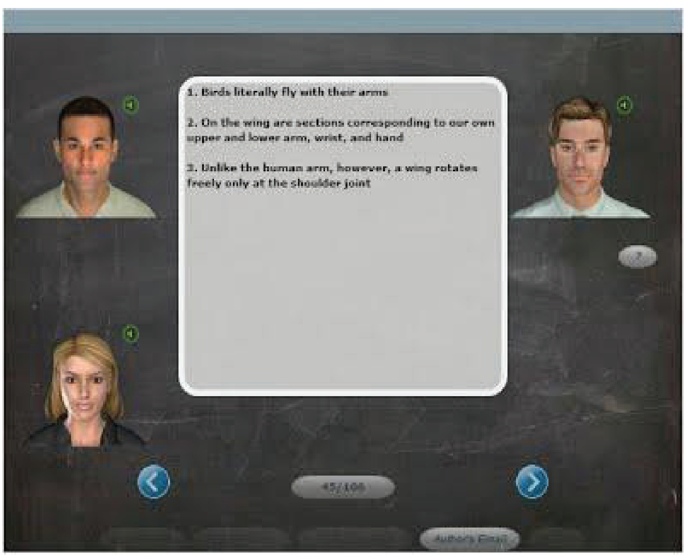

Figure 1. Example screen capture of AutoTutor LITE.

Method and Procedure

Participants were first given a demographics questionnaire asking them basic information about age, year in school and major. Following the demographics questionnaire, participants completed the Gates MacGinite Reading Comprehension test. The Gates MacGinite is designed to assess students' reading levels throughout the course of their education. Participants were given 15 minutes to complete this portion of the experiment. After the reading comprehension test, participants were given a 30 question prior knowledge questionnaire assessing general science knowledge questions in addition to history and literature (e.g., "Blood is supplied to the heart wall by the…"). Participants then completed the pretest (Earthquakes and Heart Disease counterbalanced between pretest and posttest) which consisted of two parts: 1) A paragraph broken into sentences in which learners had the opportunity to type any questions they may have about the sentence they just finished reading and 2) a multiple choice test in which they were required to answer questions about the previously read paragraph.

Participants were then randomly assigned to one of two different conditions. In the *Question Training* condition participants watched a trialogue between three animated pedagogical agents (a teacher agent and two student agents). The training begins with a brief introduction where the teacher agent discusses the importance of question asking and describes the difference between a deep and shallow question. Following the introduction, a series of science passages appear on the screen and the two student agents take turns asking questions (deep and shallow) and received feedback from the teacher agent. At predetermined points during the presentation, the participants were asked to generate their own question based on the science passages on the computer monitor and received real time feedback on their question from the experimenter. The question was rated by the experimenter on a four point scale: 1 represented a question that was "good" based on the Graesser and Person (1994) taxonomy *and* relevant to the content currently being covered in the intervention; 2 represented a question that was "good" but not relevant to the information being covered; 3 represents a question that is not good but relevant to the information currently being covered in the intervention and 4 represent a question that is bad and not relevant.

In the *Artigo* condition (control), participants were paired with an anonymous online partner and viewed various pictures on the monitor. Their job was to try to match as many words as they could with their online partner and they received points for every matching word. The participants worked on this task for 25 minutes. The justification for choosing this type of control was to have the participants work on a task that was unrelated to question asking training. In other words, to control for time on task, we wanted participants to engage in a task for a comparable amount of time that would have no impact on students' understanding on the importance of generating questions.

Following the completion of the intervention, participants completed the posttest (counterbalanced with the pretest). Finally, participants completed two tests of individual differences (i.e., the Big Five Personality Test and the Motivated Strategies for Learning Questionnaire). The Big Five personality test was designed to measure the following personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism. The Big Five personality test consisted of 45 items which were rated using a 5-point scale with 1 representing "strongly disagree" and 5 representing "strongly agree."

The MSLQ is used for assessing college students' motivational orientations and their different learning strategies during a college course. In other words, the two main constructs being measured by the MSLQ are motivation and learning strategies. The MSLQ contains two different sections and consists of 81 items (31 questions addressing motivation and 50 items addressing learning strategies). Participants answer each question using a 7-point scale where 1 represents "not at all true of me" and 7 represents "very true of me."

Table 1

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

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

RESULTS

The current study explored the possibility of teaching learners to become better question askers in a relatively short amount of time. Results revealed a significant difference in the quality of questions generated on the posttest as a function of conditions. Participants in the *Question Training* condition asked significantly more "deep" questions on the posttest than did the participants in the control condition, t (46) = 8.825, p = .000, d = 2.56. A full list of questions asking means for the pretest and posttest can be seen in Figure 2.

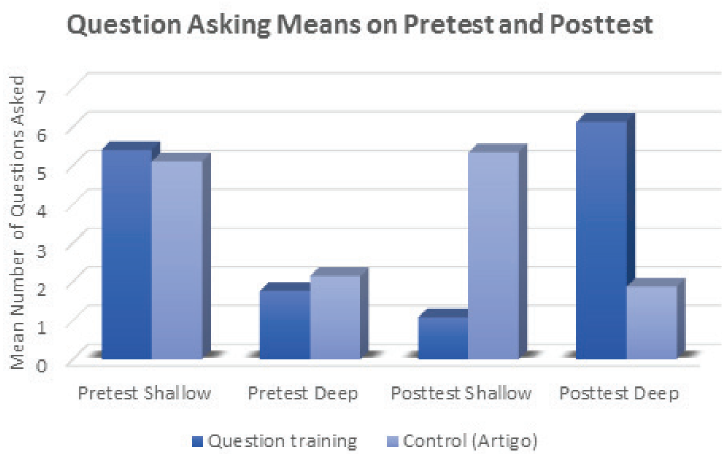

Figure 2. Question asking means on the pretest and posttest.

Because previous research has suggested that the quality of questions asked can influence learning, an analysis was performed to see if any differences existed between conditions on learning gains. Results revealed a significant difference between the two conditions on the posttest scores. More specifically, participants in the *Question Training* condition scored significantly higher than the participants in the *Artigo* condition on the posttest, F (1,45) = 10.04, p = .003, η^2 = .182. A full list of means can be seen in Figure 3.

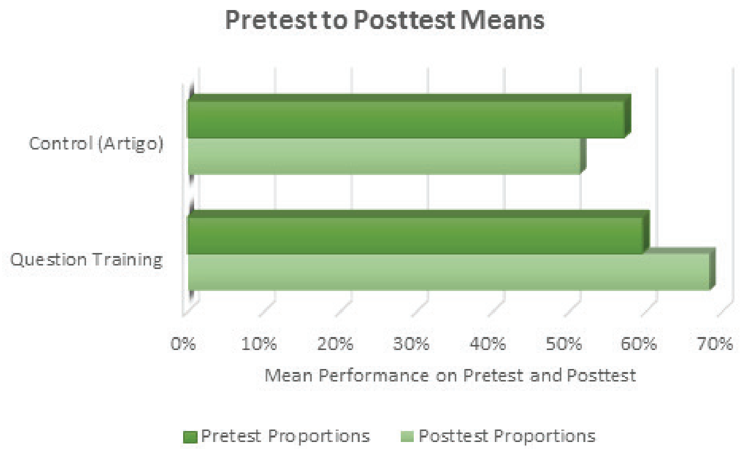

Figure 3. Pretest and posttest means across the two conditions.

In order to determine if prior knowledge had any effect on question training, the participants in the question training condition were divided into two groups: those who scored above 70% on the prior knowledge questionnaire and those who scored below 70% on the prior knowledge questionnaire. Results revealed a marginal significant interaction between prior knowledge and question training on posttest scores $F\ (1,42) = 3.541$, $p = .06$, $\eta^2 = .08$. As can be seen in Figure 4, participants in the control condition (both high knowledge and low knowledge) did not benefit from the lack of question training. Additionally, participants that entered the learning session with *high* prior knowledge did not benefit from receiving question generation training via pedagogical agents. However, those who received the question generation training that also entered the learning session with *low* prior knowledge gained the equivalent of two letter grades from pretest to posttest. Lastly, no significant effects were discovered on any of the measures of individuals differences (i.e., Big Five and MSLQ) related to learning or question asking.

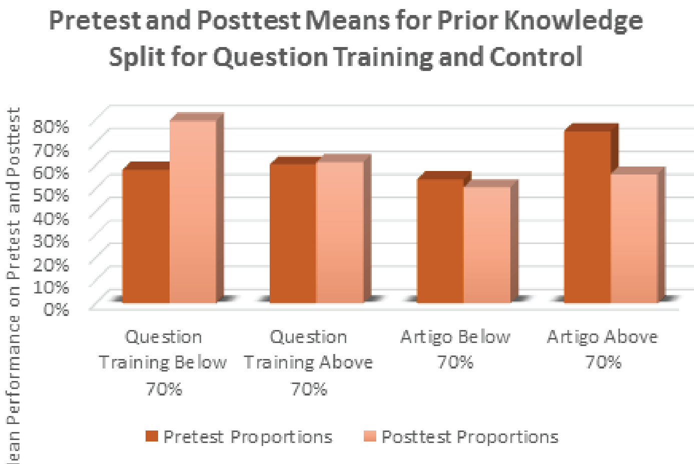

Figure 4. Pretest and posttest proportions across conditions split by prior knowledge**.**

Discussion

The current study sought to answer the question: can students be taught to become better question askers in a relatively short amount of time? The results suggest that the answer is yes. More specifically, the results seem to support the *vicarious learning theory.* Participants that simply watched a question training session among three animated pedagogical agents did ask more deep questions in addition to performing significantly better on the posttest than those in the control condition.

Further analysis seems to provide support for the *expertise reversal effect.* These data revealed question training is more beneficial for those students who enter the training with a lower level of knowledge.

How does one explain the results from the current study? It is the belief of the authors that cognitive load plays a role in these substantial results (Sweller, 1988). More specifically, cognitive load theory rests on the assumption that we as humans have a limited working memory capacity (Miller, 1956). Cognitive load can be subdivided into three different categories: intrinsic, extrinsic, and germane. Intrinsic cognitive load is a measure of the inherent difficulty of a subject area due to the number of interacting bits of information involved (e.g., learning a foreign grammar). Extrinsic cognitive load refers to invested mental effort that does not result in learning (e.g., poorly designed webpage). Germane cognitive load refers to the mental effort used to form schemas and actively integrate new incoming information with preexisting knowledge (e.g., studying for an upcoming exam).

Due to the fact that we generate questions (albeit badly) from an early age, the simple act of asking questions is a task that is low in intrinsic cognitive load. Additionally, due to the minimalistic interface of AutoTutor Lite, the current study was also low in extrinsic cognitive load. However, the task of learning how to become a "good" question asker is a task that has the potential to be high in germane cognitive load. This high germane cognitive load could have caused low knowledge learners to experience a bottleneck of information which could have been detrimental to learning. However, during the learning session, learners were able to receive explicit instructions on why question asking is good and what can be done to become a better question asker. Learners were able to view an interaction between a virtual teacher and virtual students in which students asked example questions and received feedback from the virtual teacher. This explicit instruction along with explicit examples may have served as scaffolding to the low knowledge learner which in turn could have freed up working memory capacity that the learner could have then used to form the appropriate question generation schema.

However, this exact experimental set up which turned out to be beneficial for the low knowledge learners actually worked against the learners who entered the session with high prior knowledge. This result can be explained by the *expertise reversal effect*. Because learners entered the session with preexisting knowledge (schemas), they may not need additional instructional assistance because their schemas provide full guidance. However, in the current study, instructional assistance was provided and high knowledge learners were unable to avoid this information. Because of this, there was an overlap between the instructional assistance and their existing schemas which resulted in the presentation of redundant information which required additional working memory resources which could have caused a working memory overload (Mayer, 2008).

The results from this study have immediate educational applicability. As mentioned previously, many teachers and school districts do not have the resources to provide individualized instruction to every student. The current multimedia vicarious learning is one that can be readily produced and distributed on the Internet. Access to streaming content (e.g., YouTube.com) is already ubiquitous in our everyday lives and is only expected to continue to grow at an exponential rate.

The current study seems to provide hope for those students that have a lower knowledge base from which to jump. As can be seen from the results, the students who entered the session with low prior knowledge (science, literature, and history) gained on average almost two letter grades from pretest to posttest whereas the high domain knowledge learners gained on 1% from pretest to posttest. In other words, at least in the context of the current study, low knowledge students seem to benefit the most from the question generation training which in turn causes them to surpass the high domain knowledge learners on posttest scores (79.50% compared to 61.53%).

References

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73(3)*, 277-320.

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

Azevedo, R., Guthrie, J. T., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research, 30(1-2),* 87-111.

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development, 56(1),* 45-72.

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

Bandura, A. (1962). Punishment revisited. *Journal of Consulting Psychology*, *26(4)*, 298-301.

Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives. Vol. 1: Cognitive Domain. New York: McKay.

Bourdeau, J., & Bates, A (1997). Instructional Design for Distance Learning. In R. D. Tennyson., S. Dijkstra, N. Seel, and F. Schott (Eds.)*, Instructional Design: International Perspectives*, 369-397. Hillsdale, NJ: Erlbaum.

Bringula R. P., Basa R. S., Dela Cruz C., & Rodrigo M. M. T. (2016). Effects of prior knowledge in mathematics on learner-interface interactions in a learning-by-teaching intelligent tutoring system. *Journal of Educational Computing Research, 54(4)*, 462–482.

Bugg, J. M., & McDaniel, M. A. (2012). Selective benefits of question self-generation and answering for remembering expository text. *Journal of Educational Psychology*, *104(4)*, 922-931.

Chin, C., & Brown, D. E. (2002). Student-generated questions: A meaningful aspect of learning in science. *International Journal of Science Education*, *24(5),* 521-549.

Collins, A., Brown, J.S., Larkin, K.M., Spiro, R.J., Bruce, B.C., & Brewer, W.F. (1980). Inference in text understanding. *Theoretical Issues in Reading Comprehension*. Hillsdale, NJ: Erlbaum, 385-407.

Cox, R., McKendree, J., Tobin, R., Lee, J., & Mayes, T. (1999). Vicarious learning from dialogue and discourse. *Instructional Science*, *27*, 431–458.

Craig, S. D., Chi, M. T. H., & VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of Educational Psychology*, *101(4)*, 779-789.

Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level reasoning questions effect: The role of dialog and deep-level reasoning questions during vicarious learning. *Cognition and Instruction*, 24: 563–589.

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

Draper, S. W. (2009). Catalytic assessment: understanding how MCQs and EVS can foster deep learning. *British Journal of Educational Technology*, *40*, 285–293. doi:10.1111/j.1467-8535.2008.00920.

Eva, K. W., Cunnington, J. P., Reiter, H. I., Keane, D. R., & Norman, G. R. (2004). How can I know what I don't know? Poor self-assessment in a well-defined domain. *Advances in Health Sciences Education*, *9*(3), 211-224.

Fyfe, E. R., & Rittle-Johnson, B. (2016). Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology*, *108(1)*, 82-97.

Gholson, B., & Craig, S. D. (2006). Promoting constructive activities that support vicarious learning during computer-based instruction. *Educational Psychology Review, 18(2),* 119-139.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31(1),* 104-137.

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

Graesser, A.C., Ozuru, Y., Sullins, J., McKeown, M.L., Kucan, L. (2009). What is a good question?.Threads of Coherence in Research on the Development of Reading Ability. Guilford, NY, 112-141.

Graesser, A. C., Person, N., Harter, D., & the Tutoring Research Group (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*, *12*, 257-279.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101(3),* 371-395.

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.

Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on metacomprehension accuracy. Memory & Cognition, 37(7), 1001-1013.

Hardy, J., Bates, S. P., Casey, M. M., Galloway, K. W., Galloway, R. K., Kay, A. E., Kirsop, P., & McQueen, H. A. (2014). Student-generated content: Enhancing learning through sharing multiple-choice questions. *International Journal of Science Education*, *36*(13), 2180-2194.

Heilman, M., & Smith, N. A. (2010). Tree edit models for recognizing textual entailments, paraphrases, and answers to questions. *In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, 1011-1019, Association for Computational Linguistics.

Hu, X., Cai, Z., Han, L., Craig, S. D., Wang, T., & Graesser, A. C. (2009). AutoTutor Lite. *In Proceedings of the 2009 conference on Artificial Intelligence in Education: Building learning systems that care: From knowledge representation to affective modelling*, 802, IOS Press.

Kalyuga, S., Ayers, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. The expertise reversal effect. *Educational Psychologist, 38 (1)*, 23-31.

Karabenick, S. A., & Newman, R. S. (2009). Seeking help: Generalizable self-regulatory process and social-cultural barometer. *Contemporary motivation research: From global to local perspectives*, 25-48.

Li, H., Duan, Y., Clewley, D. N., Morgan, B., Graesser, A. C., Shaffer, D. W., & Saucerman, J. (2014, January). Question asking during collaborative problem solving in an online game environment. In *Intelligent tutoring systems* (pp. 617–618). Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-07221-0_80

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*(8), 760-769.

McNamara, D.S. (1995). Effects of prior knowledge on the generation advantage: Calculators versus calculation to learn simple multiplication. *Journal of Educational Psychology, 87*, 307-318.

McNamara, D.S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55*, 51-62.

McNamra, D. S., Ozuru, Y., & Floyd, R. G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. International *Electronic Journal of Elementary Education, 4(1),* 229-257.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev., (63*), 81-97.

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

Moore, J. L., Dickson-Deane, C., & Galyen, K. (2011). e-Learning, online learning, and distance learning environments: Are they the same?. *The Internet and Higher Education*, *14(2),* 129-135.

Muller, D. & Sharma, M. (2012). Tackling misconceptions in introductory physics using multimedia presentations. *Proceedings of the Australian Conference on Science and Mathematics Education, 58-63.*

Nelson-Le Gall, S. A. (1987). Necessary and unnecessary help-seeking in children. *The Journal of genetic psychology, 148(1),* 53-62.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational psychology review, 16(4)*, 385-407.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21(2),* 267-280. doi:10.1016/j.learninstruc.2010.07.004

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

Rosenthal, T. L., & Zimmerman, B. J. (1978). *Social learning and cognition.* New York: Academic Press.

Rus, V., & Arthur, C. G. (2009). The question generation shared task and evaluation challenge. In The University of Memphis, National Science Foundation.

Speth, E. B., Momsen, J. L., Moyerbrailean, G. A., Ebert-May, D., Long, T. M., Wyse, S., & Linton, D. (2010). 1, 2, 3, 4: infusing quantitative literacy into introductory biology. CBE-Life Sciences Education, 9(3), 323-332.

Song, H. S., Kalet, A. L., & Plass, J. L. (2016). Interplay of prior knowledge, self-regulation and motivation in complex multimedia learning environments. *Journal of Computer Assisted Learning*, *32,* 31–50. doi: 10.1111/jcal.12117.

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

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

Sullins, J., Craig, S. D., & Graesser, A. C. (2010). The influence of modality on deep-reasoning questions. *International Journal of Learning Technology, 5(4),* 378-387.

Sullins, J., Howard, T., & Goza, K. (2014). Are educational shows teaching our kids how to become lifelong learners? *Journal of Educational Multimedia and Hypermedia, 23(4),* 385-400.

Sweller, J., (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Tang, I. A., Butler, D. L., Cartier, S. C., Giammarino, M., & Gagnon, F. (2006, May). Strategic help-seeking by secondary ESL students in reading contexts. In *annual meeting of the Canadian Society for Studies in Education*, Toronto, ON, Canada.

Van der Meij, H. (1988). Constraints on question-asking in classrooms. *Journal of Educational Psychology*, *3*, 401-405.

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

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45(1),* 166-183.