Instr Sci (2009) 37:487–493
DOI 10.1007/s11251-008-9069-2

Exploring the deep-level reasoning questions effect during vicarious learning among eighth to eleventh graders in the domains of computer literacy and Newtonian physics

Barry Gholson · Amy Witherspoon · Brent Morgan · Joshua K. Brittingham · Robert Coles · Arthur C. Graesser · Jeremiah Sullins · Scotty D. Craig

Received: 15 August 2007 / Accepted: 21 August 2008 / Published online: 12 September 2008

Abstract This paper tested the deep-level reasoning questions effect in the domains of computer literacy between eighth and tenth graders and Newtonian physics for ninth and eleventh graders. This effect claims that learning is facilitated when the materials are organized around questions that invite deep-reasoning. The literature indicates that vicarious learners in college student populations show greater pretest to posttest learning gains when presented with deep-level reasoning questions before each content sentence, than when deep-level questions are omitted, or when learners interact with an intelligent tutoring system. This effect holds for vicarious learners across grade levels and domains.

Keywords Vicarious learning · Deep-level reasoning questions · Questions and learning · Randomized classroom research · Physics · Computer literacy

The broad aim of this and related research is to identify conditions that support vicarious learning processes in multimedia environments and how these conditions can assist both distance learning and standard classroom instruction (Anderson et al. 1995; Chi et al. 2008; Cox et al. 1999; Craig et al. 2002; Gholson and Craig 2006; McKendree et al. 1998; McNamara et al. 2004). In these *vicarious* multimedia environments, learners are not the addresses of the educational material they are attempting to master. Instead, they can engage in only cognitive activity, neither controlling the material's source nor interacting with it in any way (Craig et al. 2006; Mayer 2001, 2002; Rummel and Spada 2005). Thus, the only activities available to learners in these vicarious environments are cognitive. The learner must actively process incoming information into coherent representations that can be integrated into existing mental models and schemas (Gholson and Craig 2006; Mayer 1997, 2001; Sweller 1988, 1999). In the next section we briefly discuss some earlier research that explored the role of deep-level reasoning questions in facilitating these cognitive activities during vicarious learning.

B. Gholson · A. Witherspoon · B. Morgan · J. K. Brittingham · R. Coles · A. C. Graesser · J. Sullins · S. D. Craig (✉)
University of Memphis, Memphis, USA
e-mail: scraig@memphis.edu

Deep-level reasoning questions and vicarious learning

Deep-level reasoning questions (hereafter deep questions) usually request one of the following three kinds of reasoning: the first is logical reasoning, which states the premise and conclusions in a syllogistic sequence. The second is causal reasoning, which articulates the states the events in a causal chain. The third is goal-orientation reasoning, which expresses the planning and goals of agents (Bloom 1956; Flavell 1963; Graesser and Person 1994; Piaget 1952, 1968). These questions ask the hearer to answer such questions as "What happens when…?", "How does the…?", and "Why was the…?". The cognitive activities required by generating and answering such deep questions are involved in comprehension (Collins et al. 1980; Rosenshine et al. 1996) and problem solving (Graesser et al. 1996; Sternberg 1987).

Craig et al. (2000) used vicarious learning procedures to show that vicarious learners presented with such deep questions embedded in course content outperform similar vicarious learners presented the same course content without such questions. This deep questions effect was soon replicated using a variety of experimental and control conditions (e.g., Craig et al. 2004, 2002; Driscoll et al. 2003).

This deep questions effect during vicarious learning was next contrasted with learning gains obtained by learners who interacted with an intelligent tutoring system (ITS) called AutoTutor. This ITS, which holds a conversation in natural language with the learner, produces learning gains of 0.6–2.1 standard deviation units when compared to various control conditions (Graesser et al. 2004, 2001; VanLehn et al. 2007). AutoTutor's script begins each topic with an overview followed by a difficult question for the tutee to answer. Sentences in an *ideal answer* to this question are decomposed into sentences containing a set of key concepts called *expectations*. Latent semantic analysis (Graesser et al. 2001; Landauer and Dumais 1997) and other semantic evaluation algorithms are used to assess the learner's progress on each topic by comparing learner contributions to the content of the expectations. The dialogue moves of AutoTutor are subsequently sensitive to content that is covered in these semantic matches so that content gaps are eventually filled by the student or tutor. Thus, the dialogue moves of AutoTutor are dynamic and adaptive.

In two experiments Craig et al. (2006) contrasted interactive learning gains of students tutored by AutoTutor with gains of vicarious learners. In Exp. 1, Craig et al. (2006) compared pretest-to-posttest gains obtained by learners in interactive tutoring sessions with gains obtained in four vicarious conditions. One vicarious condition included dialogue containing deep questions. These questions preceded each content sentence in AutoTutor's ideal answer and each expectation in the script. The other three vicarious conditions included dialogue with questions before some of the content sentences, monologue presentations of the content sentences, and videos of interactive tutoring sessions. The vicarious condition which included deep questions before each content sentence significantly outperformed the interactive condition as well as the other three vicarious conditions. In Exp. 2, Craig et al. (2006) showed the deep questions effect held when questions were embedded in monologue as well as dialogue.

Present research

Two considerations that limit the generality of the findings described in the last section are addressed in this new research. First, our earlier studies included only college students, but the new research involves eighth-to-eleventh grade students. Second, our earlier work investigated vicarious learning processes only in the domain of computer literacy, but the new research includes both the domains of computer literacy and Newtonian physics.

The present research contrasted the performance of eighth, ninth, tenth, and eleventh grade students who were tutored by AutoTutor with similar students in vicarious learning conditions. The eighth and tenth graders were asked to learn computer literacy, while the ninth and eleventh graders studied Newtonian physics. Three experimental conditions at each grade level included *interactive* tutoring sessions with AutoTutor, a vicarious-learning condition involving *monologue* presentations of each sentence in the ideal answer and each expectation (see previous section), and a vicarious learning condition which included *deep questions dialogue* in which the content of each sentence in the ideal answer and each expectation was preceded by a deep question. Based upon prior research, we predicted that vicarious learners presented with deep questions would outperform those in the monologue condition and perhaps perform as well as (learn as much) if not better than those in the interactive tutoring sessions with AutoTutor.

Method

Participants

This study included 342 participants: 74 eighth graders, 91 ninth graders, 90 tenth graders, and 87 eleventh graders. They were drawn from a public school in a mid south city. All students were tested in group settings of 11 or more in the classroom or in a laboratory at the University. Parental permission was obtained for each student prior to participation in the research.

Design, materials, and procedures

There were three experimental conditions at each of the four grade levels. The interactive condition involved participants interacting directly with an intelligent tutoring system by carrying on a dialogue with AutoTutor (112 participants). The monologue vicarious condition involved a virtual tutor asking one broad question at the outset of each topic and following this question with the sentences included in the ideal answer and expectations on that topic (116 participants). The dialogue condition involved a virtual agent asking a deep question presented by a second voice engine before each content sentence in the ideal answer and each sentence in the expectations (114 participants).

All participants at each grade level were randomly assigned to one of the three experimental conditions at the outset of their first session. After informed consent they were administered the Gates-McGinite reading test (MacGinitie et al. 2000), followed by a pretest. A 26 (three or four foil) multiple-choice questions (chance = 7 correct) was presented to participants asked to learn Newtonian physics. The test was developed by VanLehn et al. (2007) and was similar to the Force Concepts Inventory (Hestenes et al. 1992). A second version this test, used to evaluate learning gains, was administered at the completion of the study. On computer literacy two 24-item four-choice multiple choice tests (chance = 6) were used to evaluate pretest to posttest gains (see Craig et al. 2004, 2006). The two versions of the tests in each domain (both computer literacy and Newtonian physics) were shown to be equivalent in previous research (see Craig et al. 2004, 2006; Graesser et al. 2004; VanLehn et al. 2007). The two versions of the tests used in each domain were administered in counterbalanced order to learners in each condition. After completing the pretest they were presented with one of the three computerized conditions. Each computerized session was individually presented via a laptop computer located on the student's desk. All tutoring sessions were approximately 37 min in length. Altogether, the complete study required most of each of three classroom sessions for each participant who was tested in the schools (a total of about 130 min) and about 130 min for those tested in the laboratory setting.

Results and conclusions

There were no significant differences on the Gates-McGinite test among the three experimental groups within any of the four grade levels. There were also no significant differences between computer literacy and Newtonian physics on the pretests, $F(1,339) = 1.10$, $p = 0.90$. There were no significant differences at pretest among the three experimental groups (interactive, monologue, and dialogue) at any of the four grade levels (largest $F = 0.83$, $p = 0.42$). There were also no significant effects of grade level on the pretest (see Table 1).

An analysis of covariance was performed on posttest scores in a 3 (experimental condition: dialogue versus interactive versus monologue) $\times$ 4 (grade: 8 vs. 9 vs. 10 vs. 11), using pretest scores as covariates. There were significant effects of experimental condition, $F(2,329) = 6.35$, $p < 0.05$, and grade, $F(3,329) = 43.94$, $p < 0.001$. The interaction did not approach significance ($p > 0.80$). The effect of grade reflected significantly higher scores among both ninth and eleventh graders than among both eighth ($p < 0.01$) and tenth graders ($p < 0.01$). Neither eight and tenth graders nor ninth and eleventh graders differed significantly from each other ($p > 0.06$ in each case). It will be recalled that eight and tenth

Table 1 Means, standard deviations, adjusted posttests, and pre-post Cohen's *d* effect size for pretest and posttest data, by grade level and condition

Condition	Pretest		Posttest		Adjusted posttest	Cohen's *d*
	M	SD	M	SD		
Eighth						
Dialogue	0.21	0.07	0.25	0.09	0.25	0.43
Interactive	0.22	0.08	0.24	0.08	0.24	0.28
Monologue	0.22	0.08	0.23	0.09	0.23	0.03
Ninth						
Dialogue	0.31	0.10	0.42	0.12	0.43	0.96
Interactive	0.32	0.11	0.35	0.16	0.36	0.28
Monologue	0.33	0.11	0.38	0.14	0.38	0.43
Tenth						
Dialogue	0.25	0.08	0.30	0.10	0.30	0.57
Interactive	0.22	0.09	0.27	0.10	0.27	0.55
Monologue	0.24	0.08	0.27	0.11	0.27	0.28
Eleventh						
Dialogue	0.30	0.12	0.43	0.15	0.43	0.94
Interactive	0.32	0.10	0.39	0.14	0.39	0.61
Monologue	0.31	0.11	0.41	0.16	0.41	0.71

graders learned computer literacy, while ninth and eleventh graders learned physics. A comparison of computer literacy with physics was significant, $T_{(340)} = 3.34$, $p < 0.001$.

The result of most interest in the current study was the effect of experimental condition (see Table 1), which was significant, as indicated in the previous paragraph. Students in the dialogue condition, with a learning gain from pretest to posttest of 31%, learned significantly more than those in both the interactive ($p < 0.04$), with 14% gains and the monologue conditions ($p < 0.05$), with gains of 17%. The interactive and monologue conditions did not differ significantly from each other. These data indicate that the deep questions effect during vicarious learning is more robust than was previously demonstrated, in that it had been obtained only among college students in the domain of computer literacy (see Craig et al. 2006; Gholson and Craig 2006).

Before concluding, we highlight some of the limitations in our current knowledge and some issues that may need to be addressed by those concerned with distance learning, and computer-based courses in general (Anderson et al. 1995; Graesser et al. 2001; McNamara et al. 2004). First, to date, our research on the role of deep questions during vicarious learning has involved only the domains of computer literacy and Newtonian Physics. In what other domains does it hold? Second, are the deep question categories (i.e., causal antecedent, causal consequent, comparison, enablement, instrumental procedural, interpretation) that were used (see Graesser and Person 1994, pp. 110–111) equal in supporting knowledge construction during vicarious learning? Future research should answer this question, identify the features of these categories that support construction, and determine what other features of discourse support these processes.

In conclusion, we note that the deep questions effect is robust in that it generalizes (a) from college students to middle and high school students, and (b) from the relatively declarative domain of computer literacy to the more causal domain of physics. While, as indicated directly above, much further research is needed, the current findings do have some practical implications. First, embedding deep questions into course content presented via computers used in the classroom or in distance learning environments may produce real benefits for learner at little cost, in terms of time or money. Second, the finding that deep questions embedded in course content can perform at least as well (or outperform) an ITS, has clear implications for return on investment, in that it is much more costly to produce an ITS than it is to simply embed deep questions in a computerized script.

Acknowledgments This research was supported by the Institute for Education Sciences (IES) Grant R305H0R0169. The tutoring Research Group (TRG) at the University of Memphis is an interdisciplinary research team composed of approximately 35 researchers from psychology, computer science, physics and education (http://www.autotutor.org). The research on AutoTutor was supported by National Science Foundation (SBR 9720314, REC 0106965, ITR 0325428) and the DOD Multidisciplinary University Research Initiative (MURI) administered by ONR under grant N00014-00-1-0106. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of IES (DOE), DOD, ONR, or NSF.

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences, 4*, 167–207. doi:10.1207/s15327809jls0402_2.

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

Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science, 32*, 301–341.

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

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., Driscoll, D., & Gholson, B. (2004). Constructing knowledge from dialogue in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. *Journal of Educational Multimedia and Hypermedia, 13*, 163–183.

Craig, S. D., Gholson, B., & Driscoll, D. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology, 94*, 428–434. doi:10.1037/0022-0663.94.2.428.

Craig, S. D., Gholson, B., Ventura, M., Graesser, A. C., & Tutoring Research Group. (2000). Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning. *International Journal of Artificial Intelligence in Education, 11*, 242–253.

Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level reasoning effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction, 24*, 565–591. doi:10.1207/s1532690xci2404_4.

Driscoll, D., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. C. (2003). Vicarious learning: Effects of overhearing dialogue and monologue-like discourse in a virtual tutoring session. *Journal of Educational Computing Research, 29*, 431–450. doi:10.2190/Q8CM-FH7L-6HJU-DT9W.

Flavell, J. H. (1963). *The developmental psychology of Jean Piaget*. New York: Van Nostrand Reinhold.

Gholson, B., & Craig, S. D. (2006). Promoting constructive activities that support learning during computer-based instruction. *Educational Psychology Review, 18*, 119–139. doi:10.1007/s10648-006-9006-3.

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

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology, 10*, S17–S32. doi:10.1002/(SICI)1099–0720(199611)10:7<17::AID-ACP435>3.0.CO;2-7.

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

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

Hestenes, G., Wells, M., & Swacjgannerm, G. (1992). Force concept inventory. *The Physics Teacher, 30*, 141–158. doi:10.1119/1.2343497.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240. doi:10.1037/0033-295X.104.2.211.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2000). *Gates-MacGinitie reading tests* (4th ed.). Itasca, IL: Riverside.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*, 1–19. doi:10.1207/s15326985ep3201_1.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2002). Using illustrations to promote constructivist learning from science text. In J. Otero, J. A. Leon & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 333–356). Mahwah, NJ: Lawrence Erlbaum Associates.

McKendree, J., Stenning, K., Mayes, T., Lee, J., & Cox, R. (1998). Why observing a dialogue may benefit learning. *Journal of Computer Assisted Learning, 14*, 110–119. doi:10.1046/j.1365-2729.1998.1420110.x.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iStart: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers, 36*, 222–233.

Piaget, J. (1952). *The child's conception of number*. London: Routledge and Kegan Paul.

Piaget, J. (1968). *Six psychological studies*. New York: Vintage Books.

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences, 14*, 201–241. doi:10.1207/s15327809jls1402_2.

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

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

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

Sweller, J. (1999). *Instructional design in technical areas*. Melbourne: Australian Council for Educational Research.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). Natural language tutoring: A comparison of human tutors, computer tutors and text. *Cognitive Science, 31*, 3–62.