Self-explaining in the Classroom: Learning Curve Evidence

Robert G.M. Hausmann (bobhaus@pitt.edu)
Department of Psychology and the Learning Research and Development Center
University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260

Kurt VanLehn (vanlehn@cs.pitt.edu)
Department of Computer Science and the Learning Research and Development Center
University of Pittsburgh, 3939 O'Hara Street, Pittsburgh, PA 15260

Abstract

Research conducted in the laboratory and classroom has repeatedly found that self-explaining is a useful, self-directed learning strategy. Although the self-explanation effect has been replicated several times, the sources for its effectiveness are still under investigation. The present study attempts to address the question: Why does self-explaining work? Two alternative proposals are contrasted. The *content account* proposes that self-explaining is effective because of the additional information to which the learner is exposed. Alternatively, the *generation account* suggests that it is the activity of producing an explanation that is effective. The evidence, taken from learning curves collected in the classroom, predominantly supports the generation account of self-explanation, which highlights the benefit of actively processing the learning material, instead of simply attending to it.

Keywords: Self-explanation; paraphrasing; physics education research; study strategies.

To help smooth the transition from novice to expert-like performance, the cognitive and learning sciences have focused upon learning strategies that have proven to be effective across several different content domains. Among these domain-independent learning strategies is *self-explaining*, which is defined as the sense-making process that an individual uses to gain a deeper understanding of some instructional material, including textbooks, worked-out examples, diagrams, and other multimedia materials (Roy & Chi, 2005).

Self-explaining has consistently been shown to be effective in producing robust learning gains in the laboratory (Butcher, 2006; Chi, Bassok, Lewis, Reimann, & Glaser, 1989), in the classroom (McNamara, Levinstein, & Boonthum, 2004), with prompting from humans (Chi, DeLeeuw, Chiu, & LaVancher, 1994) and computers (Aleven & Koedinger, 2002; Conati & VanLehn, 2000; Hausmann & Chi, 2002).

Given the utility of self-explaining, it is important to understand the mechanisms that underlie its effect. These mechanisms, however, are still being investigated.

Why does self-explanation work?

One of the open questions with respect to the self-explanation effect is why strong learning gains are observed across several disciplines and learning contexts. In other words, why does self-explanation work? Two potential explanations will be addressed in the paper that follows. The first explanation asserts that the differences in the *content* are responsible for the increased learning gains. That is, self-explaining generates additional information that is not present in the instructional materials. Alternatively, learning from self-explaining might arise from the *activity* of producing the explanations, which is independent of the content that is produced. To explicitly contrast these two explanations, let us provide names for the hypotheses. The first is the content-account and the second is the generation-account of self-explaining.

The content-account of self-explaining. One of the outcomes of self-explaining is the inference of new knowledge. The quality of that knowledge, however, is highly variable (Renkl, 1997). In fact, explanations can vary along a continuum, from high- to low-quality. For example, consider the contrast in content between an *instructional explanation* (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001), which might be considered a high-quality explanation, and a *student-produced* explanation (Chi, 2000) (see Table 1). Both excerpts were taken from the domain of the human circulatory system.

Table 1. A contrast between instructional and student-produced explanations

Instructional Explanation	Student-produced Explanation
S7: *The right side pumps blood to the lungs and the left side pumps blood to other parts of the body.*	S32: *The muscles of the right ventricle contract and force blood through the right semi-lunar valve and into the vessels leading to the lungs.*
"That's right, the right side receives the blood, pumps it into the lungs, the lungs bring it back into the left side and the left side pumps it to the left side through the aorta."	(pause) "Um, I mean, I guess I understand now. I just, I can't think. I don't know, but kind of a muscle contraction that pushed the blood, um, through the valve and into vessels, but I don't know."

Note: the text in *italics* is the current sentence of the text.

One of the notable differences between the two explanations is the completeness and coherence. The instructional explanation, generated by a nursing student tutoring a student, is more coherent and complete than the student-produced explanation. Furthermore, the student's explanation contains speech disfluencies, sense-making statements, and meta-cognitive comments.

Given the differences between the two explanations, the content-account of self-explaining predicts that the quality of an explanation will determine the overall learning. Thus, if an instructional explanation is of a higher quality than that produced by a student, then the instructional explanation will be more effective because it is more coherent and complete.

The generation-account of self-explaining. In contrast, the generation-account of self-explaining suggests that it is important for the student to actively produce the explanation. During self-explaining, the student is engaged in an active learning process, which includes accessing prior knowledge from long-term memory, using common-sense reasoning, employing sense-making strategies, and doing so from their own background knowledge. Therefore, there may be something special about the *activity* of explaining that is important for learning.

Indirect evidence for the generation-account can be found in the literature on human memory. One of the consistent findings is the *generation effect*, which states that items produced by an individual are more likely to be recalled or recognized at a later point in time (Jacoby, 1978; Slamecka & Graf, 1978). This robust memory effect has been thoroughly tested on simple verbal items (McNamara & Healy, 2000), as well as more complex stimuli, including single sentences (Kane & Anderson, 1978), trivia questions (deWinstanley, 1995), and even conceptual material (Foos, Mora, & Tkacz, 1994). However, one of the limitations of the generation effect is that the information being tested already resides in long-term memory. Therefore, it is an open question if the generation effect can generalize to more complex domains, such as procedural or conceptual learning, where the information is new to the student.

Disaggregating content from generation. There have only been a few empirical studies that attempt to disaggregate the effects of content from the activity of explaining (Brown & Kane, 1988; Schworm & Renkl, 2002). An exemplary case can be found in a study by Lovett (1992), in which she crossed the source of the solution to permutation and combination problems (subject vs. experimenter) with the source of the explanation for the solution (subject vs. experimenter). The experimenter-subject condition was analogous to experimental materials found in a self-explanation experiment because the examples were incomplete, and the experimenter-experimenter condition was analogous to studying a complete, worked-out example. Lovett found that the subject-subject condition and the experimenter-experimenter condition demonstrated the best

performance, especially on far-transfer items. Lovett's interpretation was that the subject-subject condition was effective because the students were actively engaged and the experimenter-experimenter condition was effective because it contained higher quality explanations than those generated by students. To better understand the pattern of results, Lovett analyzed the protocol data and found that for the participants who generated the key inferences, their learning gains were the same as participants who read the corresponding concepts.

Brown and Kane (1988) found a similar pattern of results. They demonstrated that explanations provided by children (4-7 years old), either spontaneously or in response to prompting, were much more effective at promoting transfer than those provided by the experimenter. In particular, students were first told a story about mimicry. Some students were then told, "Some animals try to look like a scary animal so they won't get eaten." Other students were first asked, "Why would a furry caterpillar want to look like a snake?" and if that didn't elicit an explanation, they were asked, "What could the furry caterpillar do to stop the big birds from eating him?" Most students got the question right, and if they did, 85% were able to answer a similar question about two new stories. If they were told the rule, then only 45% were able to answer a similar question about the new stories. However, the students who were told the rule may not have paid much attention to it, according to Brown and Kane.

Given that most studies confound the content with the activity of explaining, we conducted a study in which the two accounts of self-explaining were contrasted, which make the following predictions:

- *Content*: student-produced explanation = author-provided explanation > no explanation
- *Generation*: student-produced explanation > author-provided explanation = no explanation

Method

Participants and Design

One-hundred and four students, recruited from five sections of a second-semester, calculus-based physics course taught at the U.S. Naval Academy, were given course credit for their participation ($N = 104$).

The experiment was a 2 x 2 between-factors design, which crossed: Activity (self-explaining vs. paraphrasing) and Content (complete vs. incomplete). Participants were block-randomized into one of the four experimental conditions: paraphrase complete examples ($n = 26$), paraphrase incomplete examples ($n = 23$), self-explain complete examples ($n = 27$), and self-explain incomplete examples ($n = 28$). The block-randomization technique was used to ensure that the groups were equal according to GPA, Physics I grade, and exposure to the Andes homework tutor. There were no statistically reliable differences between these variables for the four conditions (all $ps > .30$).

Materials

The domain covered during the present experiment was electrodynamics, with an emphasis on the forces acting on a charged particle due to the presence of an electric field. The training materials[1] (i.e., problems, examples, and prompts) were developed in association with one of the LearnLab[2] instructors and two other physicists.

Procedure

The data were collected in the Physics LearnLab, which is a course that was designed to conduct rigorous, *in vivo* experiments on issues related to robust learning. There were two reasons for collecting the data in the LearnLab, as opposed to the laboratory. First, the realism of the classroom increases the generalizability of the results, without sacrificing randomization or some of the control over extraneous variables. Second, the LearnLab provided a facility for collecting micro-genetic log-file and verbal data from the students while they learned from examples and solved problems.

The students were introduced to the experiment and shown instructions for their learning Activity (either self-explaining or paraphrasing). Students were then prompted to solve the first problem, which was intended as a warm-up problem to acquaint the students with the Andes[3] interface. Andes is an intelligent tutoring system, which was created to help students solve homework problems from the first two semesters of introductory physics (VanLehn et al., 2005). After solving the first problem, the students then studied the first example. This process, alternating between solving problems and studying examples (Trafton & Reiser, 1993), repeated for three cycles so that by the end of the training, four problems were solved and three examples were studied.

While the students were studying the examples, they were prompted to either self-explain or paraphrase at the end of each segment. To capture their verbalizations, each student was outfitted with a pair of headphones equipped with a close-talk, noise-cancelling microphone. In addition to audio, all of the on-screen activity was recorded using a facility built into the Andes interface. The following data-streams were created for each student: 1. an audio track of their verbalizations; 2. a video of their on-screen activities; and 3. a log file of each action in the Andes interface. In addition to these three data sources, log files from the assigned homework problems, solved with Andes, were made available to the researchers.

Knowledge Components and Learning Curves

One of the assumptions made by the LearnLab is that knowledge is partially decomposable into individual components. Knowledge components (KCs) are abstract units of knowledge, which include concepts, principles,

rules, declarative knowledge, and schemata. Similar assumptions appear in other computational models of cognition, including production rules in ACT-R (Anderson & Lebiere, 1998) and chunks in SOAR (Newell, 1990).

The advantage of assuming knowledge is partially decomposable is that it allows researchers to track learning of individual knowledge components over time. This fined-grained analysis can be represented as learning curves, which plot an assistance score against the opportunity to apply that particular knowledge component. An assistance score is defined as the sum of all the errors and requests for help on that particular knowledge component (see Cen, Koedinger, & Junker, 2006 for an example). The assumption is that as students learn, the number of errors will decrease, as well as their need for help. Thus, a decrease in assistance scores reflects a fine-grained measure of learning over time.

Analyses and Results

At the level of the condition, that is, collapsing across individual problems and all knowledge components, there was a main effect for Activity, with the self-explaining condition demonstrating lower assistance scores than the students in the paraphrasing condition, $F(1, 73) = 6.19$, $p = .02$, $\eta^2 = .08$. This result suggests that the students' problem-solving performance was enhanced by the prompts to self-explain, which replicates prior laboratory results. It also lends support to the generation-account of self-explaining. How does problem-solving performance look when we use a finer grained analysis of students' performance as it unfolds over time?

To address this question, knowledge components that were necessary to solve all four problems were selected. Four knowledge components met this criterion, which included: KC1. applying the definition of the electric field ($\mathbf{F} = q\mathbf{E}$); KC2. drawing an electric-field (E-field) vector; KC3. drawing an electric-force vector; and KC4. defining the charge on a particle. The most important knowledge component was applying the definition of the electric field because it was the main principle taught in the chapter on electric fields.

To measure performance on these knowledge components, we conducted separate 2x2x3 repeated-measures ANOVA on the assistance scores for each of the four knowledge components, with Activity and Content as between-subjects factors and Problem as a within-subjects factor. Because the first problem was intended to familiarize the students with the Andes interface, analyses were restricted to the other three problems.

Because a repeated-measures ANOVA requires a value for each observation for all factors, the data were reduced in another way. Although all of the students in the sample were given three opportunities to apply each knowledge component, not every student was able to complete all three problems because each successive problem was more complex than its predecessor; therefore, the data were restricted to those who applied each knowledge component

across all three problems. The number of students who did not fit this requirement did not differ between conditions.

KC1. Applying the definition of the electric field. For the principle knowledge component, the assistance score decreased for all of the experimental conditions (see Figure 1). There were no significant main effects or an interaction. However, a post-hoc comparison between the incomplete self-explanation ($M = 2.56$, $SD = .58$) and complete paraphrase ($M = 4.23$, $SD = .64$) condition revealed a marginal difference, $F(1, 58) = 3.73$, $p = .06$, $\eta_p^2 = .06$. The difference was most pronounced for the first problem, $F(1, 58) = 4.79$, $p = .03$, $\eta_p^2 = .08$.

This pair-wise difference is consistent with previous literature that shows a strong correlation between self-explanation and learning, but not between paraphrasing and learning (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Hausmann & Chi, 2002). Unfortunately, the difference between the incomplete self-explanation and complete paraphrase confounds both the activity of generation with the content of the explanations; therefore, the evidence is equivocal for both the generation and content accounts of self-explaining.
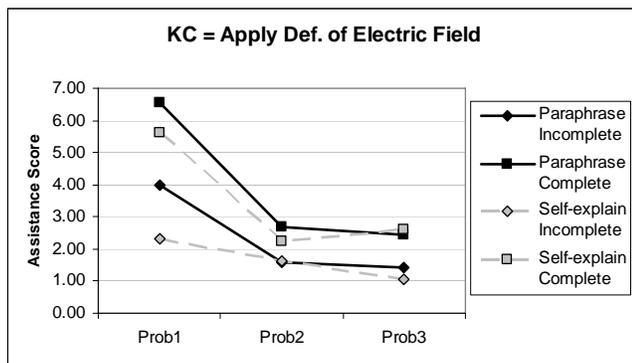

Figure 1: Assistance score per opportunity to apply the definition of an electric field ($\mathbf{F} = q\mathbf{E}$).

KC2. Drawing the electric-force vector. The pattern of results for the second knowledge component (i.e., drawing an electric-force vector) was similar (see Figure 2). There was a between-subjects main effect for Activity, with the self-explaining conditions demonstrating lower assistance scores than the paraphrasing conditions, $F(1,54) = 4.36$, $p = 0.04$, $\eta_p^2 = .07$. Unlike the first knowledge component, this result was unequivocal and consistent with the generation account.

A post-hoc comparison between the incomplete self-explanation ($M = 3.47$, $SD = 3.79$) and complete paraphrase ($M = 8.60$, $SD = 10.15$) condition also revealed a reliable difference, $F(1, 54) = 4.93$, $p = .03$, $\eta_p^2 = .08$. The differences per Problem were strongest for the first ($\eta_p^2 = .06$) and last problems ($\eta_p^2 = .05$).
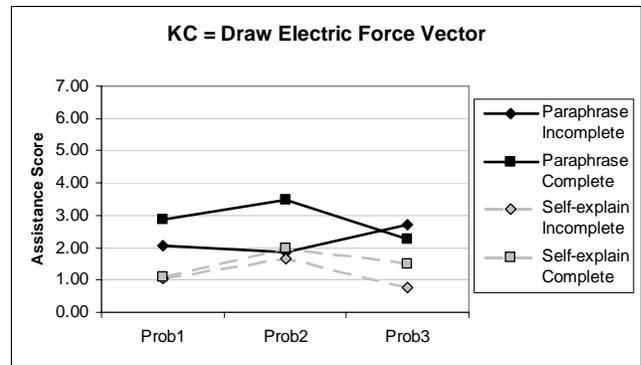

Figure 2: Assistance score per opportunity to draw the electric force vector.

KC3. Drawing an electric-field vector. A similar pattern of results was found for drawing the E-field vector. Although not statistically reliable, there was a trend for a main effect for Activity (see Figure 3), with self-explainers committing fewer errors and asking for fewer hints than the paraphrasing condition, $F(1, 71) = 2.10$, $p = 0.15$, $\eta_p^2 = .03$. This pattern of results partially supports the generation account.

A post-hoc comparison between the incomplete self-explanation and complete paraphrase condition revealed a marginally significant difference, $F(1, 54) = 3.27$, $p = .08$, $\eta_p^2 = .04$. Unlike the first two KCs, there was a trend for the incomplete self-explainers demonstrating lower assistance scores; however, none of the differences between Problems for these two conditions were reliably different.
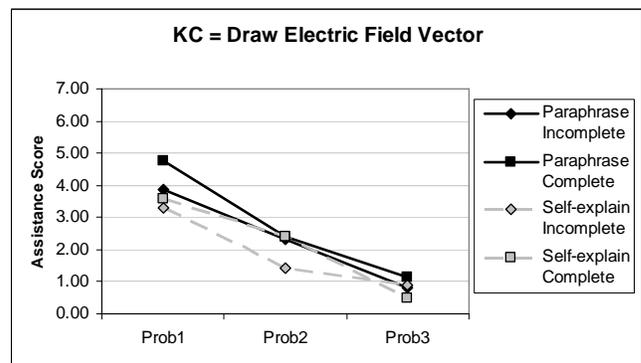

Figure 3: Assistance score per opportunity to draw an electric field vector.

KC4. Defining the charge on a particle. Finally, for the last knowledge component, defining the charge on a particle, there were neither main effects nor an interaction between conditions (all Fs < 1). The assistance scores were, however, lower for KC4 than the average assistance score for all the other KCs, $F(1,32) = 76.59$, $p < 0.001$, $\eta_p^2 = .70$.

The reason why this knowledge component was unaffected by the experimental manipulations was because

the students were committing very few errors, even on the first opportunity ($M = .59$, $SD = .17$). Defining the charge on a particle is extremely easy because all of the information is given in the problem statement. Therefore, there is little surprise why there was a large effect between KCs, yet no difference between conditions.

Discussion

The analyses of problem solving and learning at the knowledge-component level suggest a few conclusions. First, it appears that two out of the four knowledge components were learned before the first problem was solved. Drawing the electric-force vector and defining the charge on a particle exhibited flat learning curves for each of the four experimental conditions. There are at least three explanations for flat learning curves.

One interpretation of a flat curve is that no measurable learning took place from the first to the last application of that particular knowledge component. For example, there are at least four different time points in which the students might have learned how to draw an electric-force vector. The students could have learned this knowledge component during their first semester of Physics; while drawing the E-field vector (see the third explanation below); during the warm-up problem, or while studying the first example. Unfortunately, the present analyses do not provide evidence to discriminate among any of the aforementioned sources; however, forthcoming analyses of the verbal protocols may help to exclude some of the hypothesized sources.

A second explanation for a flat curve is the result of an incorrect knowledge decomposition used to define the force-vector knowledge component (Corbett, McLaughlin, & Scarpinatto, 2000). For instance, when drawing a vector in Andes, several variables need to be specified, including the *body* of interest, the *type* of force, the *angle* in which the force is acting, as well as the *time* interval. Future analyses will decompose the force vectors into their constituent sub-components to see if a monotonic decrease in assistance scores emerges.

The third explanation is that a flat learning curve may be the result of the interaction between individual knowledge components. For instance, drawing the force vector before the E-field vector may reduce the errors on the E-field vector because all of the reasoning, and therefore errors and hint-requests, will be associated with the force vector.

To further illustrate this possibility, a few details of the materials need to be considered. The first problem states, "*The force on the electron due to the electric field exactly cancels its weight near the Earth's surface.*" If the student attempts to draw the force vector first, then she must traverse the following chain of reasoning (see Table 2):

Table 2. Chain of reasoning for the force vector

IF	there is no net force,
AND	the weight is acting downward
THEN	the direction of the electrical force is upward.

After the student draws the force vector, she can then consider the E-field vector. Drawing the direction of the E-field becomes a straight-forward chain of reasoning, once the direction of the force vector is known (see Table 3):

Table 3. Chain of reasoning for the E-field vector

IF	the electric force is upward,
AND	the charge on the particle is negative,
THEN	following the vector equation $\mathbf{F}=q\mathbf{E}$, the E-field is in the opposite direction (i.e., downward).

Alternatively, when the E-field vector is considered first, then the reasoning from the force KC must be nested within the first part of the chain of reasoning of the E-field (i.e., all of Table 2 is nested in the "IF" clause of Table 3). Therefore, the errors and hint requests would be associated with drawing the E-field vector because that is the element that Andes has identified as the current focus of the student's problem-solving activity.

The last set of results that requires an explanation is the inconsistently reliable main-effects across all four knowledge components. Whereas the main effect for Activity for KC2 was significant, the same main effect for KC3 was marginal, and KC1 and KC4 were not significant. One reason why this might be the case could be due to the difficulty of learning the individual knowledge components. Future analyses of the homework log files from a different semester of students will provide an independent assessment of the difficulty for each knowledge component.

In conclusion, the analyses at the knowledge-component level partially reflect the pattern of results taken at a higher grain size (Hausmann & VanLehn, 2007). Specifically, there was a learning advantage for the students who were instructed to self-explain an incomplete example over those who were asked to paraphrase the instructional explanation. This lends support to the generation account of self-explanation, and it underscores the importance of actively engaging students in the learning material, instead of requiring them to simply attending to it. Future research will include analysis of the verbal protocols to code for the correctness and depth of the inferences, explanations, and justifications generated by the students.

Acknowledgments

References

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explain with a computer-based Cognitive Tutor. *Cognitive Science, 26*, 147-179.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Brown, A. L., & Kane, M. J. (1988). Preschool children can learn to transfer: Learning to learn and learning from example. *Cognitive Psychology, 20*(4), 493-523.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology 98*(1), 182-197.

Cen, H., Koedinger, K., & Junker, B. (2006). *Learning Factors Analysis - A General Method for Cognitive Model Evaluation and Improvement.* Paper presented at the 8th International Conference on Intelligent Tutoring Systems, Taiwan.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161-238). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13*, 145-182.

Chi, M. T. H., DeLeeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*(4), 471-533.

Conati, C., & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education, 11*, 398-415.

Corbett, A. T., McLaughlin, M., & Scarpinatto, K. C. (2000). Modeling student knowledge: Cognitive tutors in high school and college. *User Modeling and User-Adapted Interaction, 10*, 81-108.

deWinstanley, P. A. (1995). A generation effect can be found during naturalistic learning. *Psychological Bulletin & Review, 2*(4), 538-541.

Foos, P. W., Mora, J. J., & Tkacz, S. (1994). Student study techniques and the generation effect. *Journal of Educational Psychology, 86*(4), 567-576.

Hausmann, R. G. M., & Chi, M. T. H. (2002). Can a computer interface support self-explaining? *Cognitive Technology, 7*(1), 4-14.

Hausmann, R. G. M., & VanLehn, K. (2007). *Explaining self-explaining: A contrast between content and generation.* Paper presented at the 13th International Conference on Artificial Intelligence in Education, Marina Del Rey, CA.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution *Journal of Verbal Learning and Verbal Behavior, 17*(6), 649-667.

Kane, J. H., & Anderson, R. C. (1978). Depth of processing and interference effects in the learning and remembering of sentences. *Journal of Educational Psychology, 70*(4), Journal of Educational Psychology.

Lovett, M. C. (1992). Learning by problem solving versus by examples: The benefits of generating and receiving information. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 956-961). Hillsdale, NJ: Erlbaum.

McNamara, D. S., & Healy, A. F. (2000). A procedural explanation of the generation effect for simple and difficult multiplication problems and answers. *Journal of Memory and Language, 43*, 652-679.

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

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*(1), 1-29.

Roy, M., & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271-286). Cambridge: Cambridge University Press.

Schworm, S., & Renkl, A. (2002). Learning by solved example problems: Instructional explanations reduce self-explanation activity. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 816-821). Mahwah, NJ: Erlbaum.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592-604.

Trafton, J. G., & Reiser, B. J. (1993). The contributions of studying examples and solving problems to skill acquisition. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society* (pp. 1017-1022). Hillsdale, NJ: Erlbaum.

VanLehn, K., Lynch, C., Schultz, K., Shapiro, J. A., Shelby, R., Taylor, L., et al. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education, 15*(3), 147-204.