

The Design of Self-explanation Prompts: The Fit Hypothesis

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

Timothy J. Nokes (nokes@pitt.edu)
Learning Research and Development Center
University of Pittsburgh, 3939 O'Hara Street
Pittsburgh, PA 15260

Kurt VanLehn (Kurt.Vanlehn@asu.edu)
Department of Computer Science and Engineering
Arizona State University, P.O. Box 878809
Tempe, AZ 85287 – 8809

Sophia Gershman (sgershman@whrhs.org)
Watchung Hills Regional High School
108 Stirling Road
Warren, NJ 07059

Abstract

Cognitive science principles should have implications for the design of effective learning environments. The *self-explanation principle* was chosen for the current project because it has developed significantly over the past few years. Early formulations suggested that self-explanation facilitated *inference generation* to supply missing information about a concept or target skill, whereas later work suggested that self-explanation facilitated *mental-model revision* (Chi, 2000). To better understand the complex interaction between prior knowledge, cognitive processing, and changes to a learner's representation, three different types of self-explanation prompts were designed and tested in the domain of physics problem solving. The results suggest that prompts designed to focus on problem-solving steps led to a sustained level of engagement with the examples and a reduction in the number of hints needed to solve the physics problems.

Keywords: self-explanation; prompting; worked-out examples; intelligent tutoring systems.

Introduction

In many formal domains, such as mathematics, physics, and logic, it is not uncommon for students to learn from both studying examples and solving problems. Instructors typically require students to solve problems, which motivates students to use and study the examples. Unfortunately, however, some students study examples using shallow cognitive strategies, such as paraphrasing (Hausmann & Chi, 2002) or by matching surface features with equations (Ross & Kilbane, 1997; VanLehn, 1998). If students rely on these strategies, then they may flounder when solving difficult problems (Aleven & Koedinger, 2002). In an effort to avoid shallow processing, researchers have developed several kinds of prompts that facilitate *self-explanation*, which promotes both deep processing of the examples and robust learning (Chi, DeLeeuw, Chiu, & LaVancher, 1994).

The purpose of the current study is twofold. The first goal is to conduct translational research that tests different theory-driven implementations of the self-explanation principle in an applied classroom setting. Much work on learning in the cognitive sciences has implications for educational practice; however, a large divide often separates the principles discovered in the laboratory from their implementation in classroom settings. Implementing a learning principle is not

simply a straightforward translation of interpreting a textual description of that principle for a particular problem or task. Instead, it relies upon understanding the principle and the critical factors that affect its implementation. For self-explanation, one needs to take into account the learner's prior knowledge, the target knowledge or skill to be acquired, and the structure of the domain, task, or activity. Our aim here is to begin to close this divide by taking a principled approach to testing different implementations of the self-explanation principle in a classroom setting.

Our second goal is to test the *Fit Hypothesis*, which is the idea that the efficacy of prompting is contingent on the match between the cognitive processing that the prompting elicits, modifications to the underlying representations, their relation to prior learning, and the utility of those representations for understanding and solving problems. Through pursuing these goals we hope to make progress in both applying cognitive research to real-world problems (i.e., classroom learning) as well as develop a better understanding of the basic learning mechanisms of the mind.

The paper is organized into the following sections. The next section introduces two types of self-explanation prompts: those targeted for inference generation and those targeted for mental model revision. The sections after that elaborate the Fit Hypothesis and describe an experiment to test it in the context of physics problem solving. In the last section we describe the implications of the results for understanding the relation between basic learning processes, learner factors, the learning environment, as well as the translation of cognitive principles into pedagogy.

Different Types of Self-explanation Prompts

A worked-out example, in the domain of physics, consists of a series of problem-solving steps that terminates with the problem's solution. Worked-out examples demonstrate the application of domain principles and expert solution strategies. However, examples are often incomplete with respect to the conditions under which a step applies. For instance, consider the following steps from a statics example: *Figure 1 shows an object of weight W hung by strings. Consider the knot at the junction of the three strings to be "the body."* Unfortunately for the student, the example does not explicate a reason for choosing the knot as the

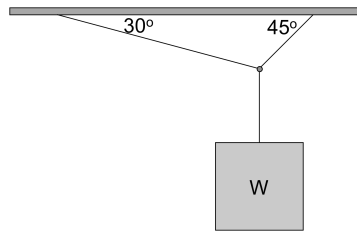


Figure 1: A figure from a worked-out example.

body (i.e., it is a choice of problem-solving convenience because the sum of the forces at the knot is zero).

In an effort to facilitate deep processing of an example, a number of self-explanation interventions have been developed, including self-explaining by referencing principles in a glossary (Aleven & Koedinger, 2002) and training vis-à-vis a tutoring system (McNamara, 2004). In the current work, we focus on two types of prompts that correspond to two hypothesized mechanisms of self-explanation, specifically targeting *gap filling* (Chi & Bassok, 1989) and *mental model revision* (Chi, 2000).

The first type of prompting, which we refer to as the *justification-based* (see Table 1a for examples) was designed around the hypothesis that self-explanation leads to learning gains because it supplies the necessary information missing from the examples. Chi and Bassok (1989) argued that good students generated self-explanation inferences to add coherence and completeness to examples that skipped steps, or did not include the application conditions for the principle or concept being applied.

Justification-based prompts have been used to successfully assist students in the early stages of learning. Conati and VanLehn (2000) constructed an intelligent tutoring system to coach students while they studied examples. The system included prompts for self-explanation that focused on the justification for taking a problem-solving step. Students who were initially learning the material demonstrated strong learning gains with a large effect size ($d = 1.07$).

The second type of prompt, which we refer to as *meta-cognitive*, was based on the hypothesis that learning occurs by revising a flawed mental model (see Table 1b for examples). Chi (2000) argued that when students read a line of text, they compare their interpretation of the text to their prior understanding. If they detect a discrepancy, then they generate a self-explanation that repairs their flawed representation.

Meta-cognitive prompts have been successfully used to facilitate learning from an expository text on the circulatory system (Chi et al., 1994). Students who were prompted to self-explain while reading the text acquired a more scientifically accurate mental model of the text than if they studied the text without prompting. Moreover, the students in the prompting condition demonstrated higher scores on the difficult items of a posttest ($d = 1.14$). The same meta-cognitive prompts have also been used to successfully

facilitate learning in other, more procedural domains, such as physics problem solving ($d = .57$, Hausmann & VanLehn, 2007).

Both justification-based and meta-cognitive prompts have been demonstrated to be effective methods for eliciting self-explanations while studying worked-out examples or reading expository texts. However, the two differ in the underlying theoretical assumptions as to whether the self-explanation primarily facilitates the generation of new knowledge (justification-based) or knowledge revision (meta-cognitive).

The Fit Hypothesis

To evaluate these two theoretical positions, we posit the *Fit Hypothesis*, which is the idea that the cognitive processes triggered by an instructional intervention must match, or fit, the particular learning situation (e.g., prior knowledge of the learner and task structure). In the case of self-explanation, we hypothesize that the justification and meta-cognitive based prompts are best suited for particular types of learning situations that depend critically on the student's prior knowledge, the relation between that knowledge and the task domain, and the structure of the task or learning activity. For example, meta-cognitive prompts are likely to facilitate deep learning when the student has prior misconceptions of the target concept, whereas justification-based prompts do not require such knowledge. The Fit Hypothesis is similar to the notion of *transfer appropriate processing* (e.g., Morris, Bransford, & Franks, 1977 – the match of cognitive processing between learning and test facilitates retrieval), but extends this concept to include a focus on the learner's prior knowledge and the structure of the task.

We can use the Fit Hypothesis to reinterpret the findings reviewed in the previous section. Chi et al. (1994) reported large effect sizes for a study that used meta-cognitive prompts to elicit self-explanation inferences whereas Hausmann & VanLehn (2007) found a much smaller effect size ($d = 1.14$ vs. $.57$) for the same meta-cognitive prompts used to encourage students to self-explain electrodynamics examples. The Fit Hypothesis would predict a reduction in the effectiveness of meta-cognitive prompting for the second study because the vast majority of the children in Chi et al. (1994) possessed incorrect mental models at the start of the experiment. Alternatively, students in Hausmann & VanLehn (2007) study were relatively new to the field of electrodynamics; therefore, it is unlikely that the students held any prior knowledge about the domain (Maloney, O'Kuma, Hieggelke, & Van Heuvelen, 2001, p. S17).

In the current study, we tested the Fit Hypothesis by examining the effectiveness of the three different types of prompts on learning from worked-out examples in physics.

The examples and problems required that students draw free-body diagrams, define vector and scalar quantities, identify relevant equations, and recognize when the problem is solved. All of these problem-solving activities might best be characterized as learning a cognitive procedural skill.

That is, learning to solve these problems requires that the student acquire a step-based, problem-solving schema.

To maximize learning from worked-out examples in this domain, students need to engage in (at least) two types of cognitive processing. First, the students should *attend* to the individual problem-solving steps illustrated in the example. Second, the students should *generate* inferences regarding the justifications for each step. The extent to which students engage in both types of processing should predict the robustness of their learning.

The Fit Hypothesis suggests that different types of prompts will be differentially beneficial because of the match between the cognitive processes and the learner's prior experience, the domain, and task constraints. First, the fit between the cognitive processes facilitated by meta-cognitive prompts and acquiring a cognitive procedural skill is low because the skill is step-based, whereas meta-cognitive prompting encourages the revision of a mental-model. Second, the fit between justification-based prompts and learning a cognitive procedural skill is high because the problem-solving activities are commensurate with attending to steps and generating inferences (i.e., reasons for taking each problem-solving step). Finally, justification-based prompting should demonstrate more robust learning than a set of prompts that merely focuses the student's attention on the problem-solving steps themselves (see Table 1c for examples).

Method

Participants

Forty-eight students were recruited from three sections of a second-semester physics course taught at an eastern high school. Volunteers were given course credit for their participation. The experiment took place in one of the open class periods, which were approximately 90 minutes in duration.

Materials

Three electrodynamics problems and two worked-out examples were designed in collaboration with the instructors of the course. Each problem-example pair was isomorphic to one another. The students attempted to solve the problem first, with the assistance of an intelligent tutoring system, and then they studied the isomorphic example. The topics covered by the problems and examples included the definition of the electric field, the weight law, Newton's second law, and several kinematics equations. Students solved the problems with the assistance of the Andes physics tutoring system, and the examples were presented as a video of the Andes screen, with an expert describing what actions were being taken. The reasons for each solution step were omitted from the examples because one of the goals of the experiment was to see if students were able to supply the missing information.

Three types of prompts were generated, one for each condition (see Table 1). The justification-based prompts

correspond to those used in Conati and VanLehn (2000). The meta-cognitive prompts were taken from a prior study that elicited self-explanations while reading an expository text (Chi et al., 1994). In addition, a third set of prompts was constructed to focus the students' attention on each step (Hausmann & Chi, 2002), but did not *require* students to generate justifications. We included this set of prompts to test whether students would spontaneously generate justifications if they were focused on explaining each step. Prompts were administered only during the example study.

Table 1. Three different types of self-explanation prompts.

a. Justification-based Prompts (Conati & VanLehn, 2000)

- What principle is being applied on this step?
- This choice is correct because...
- What is the justification for this step? Why is it correct?
- What law, definition, or rule allows one to draw that conclusion?

b. Meta-cognitive Prompts (Chi, et al., 1994)

- What new information does each step provide for you?
- How does it relate to what you've already seen?
- Does it give you a new insight into your understanding of how to solve the problems?
- Does it raise a question in your mind?

c. Step-focused Prompts (Hausmann & Chi, 2002)

- What does that step mean to you?
 - Do you have any more thoughts about that step?
 - Could you restate or summarize that step in your own words?
 - So, specifically, what else does this step tell us?
-

All of the problem solving took place within the Andes physics tutor (VanLehn et al., 2005). Andes provides several problem-solving scaffolds. First, Andes gives color-coded, instant feedback on each step. If the student enters an incorrect entry, Andes will flag the attempted step red.

Second, Andes provides the student with on-demand hints, which are graded in terms of their depth. The top-level hint directs the student's attention to a critical feature of the problem. For instance, the hint reminds the student that the body is near the earth, which should point the student to remember that gravity is acting on the body. The next level of hint teaches the student a concept or principle. Continuing the example from above (see Fig. 1), if pointing out the fact that the body is near the Earth is not enough, the next hint says, *When an object is near a planet, the planet exerts a weight force on the object.* This is a verbal description of the applicability conditions for the weight law. If this is still not enough information to take the next step, the bottom-out hint directly tells the student what to do (e.g., *draw a force on the body due to the earth of type weight*). It is important to note that the information missing in the examples can be found in the on-demand hints.

Finally, Andes also provides a list of equations that the students can reference at any time. Students are required to work with symbolic expressions while solving a problem. Once the symbolic expressions are clearly written, Andes provides the students with the algebraic solution. That is, Andes handles all of the mathematical operations of

substituting values into the variables and solves the system of equations.

Design and Procedure

The experiment was a mixed design with participants randomly assigned to one of three experimental conditions: justification-based prompting ($n = 16$), meta-cognitive prompting ($n = 15$), and step-focused prompting ($n = 17$). There were two within-subjects factors (i.e., Examples and Problems), which are introduced separately in the Results section. Students were recruited from three high-school physics courses, and the instructors explained to them that they were going to be solving problems with Andes as part of their classroom exercises.

On the day of the experiment, the students logged into Andes, and they read a short set of instructions explaining that they would be studying video-based examples. The prompts were shown to the students to orient them to what would be expected of them later. Then they solved the first problem with Andes. Afterwards, they studied the first example, which was broken down into ten steps. At the conclusion of each step, the student was prompted to self-explain the example. Below the playback screen were the same four prompts that were displayed during the introduction. In addition, talk-aloud instructions were given to the students, and the prompts were always available for the students to reference. The teacher instructed them to select and answer at least one question per step.

After completing all of the example steps, the students then solved the next problem. This cycle of solving problems and studying examples continued until the student completed all of the materials, or until the class period ended. Each of the three conditions were equally likely to finish all three problems, $\chi^2(2, N = 48) = 2.14, ns$.

The order of the problems and examples were fixed for all of the students. They all solved a problem first and then studied an isomorphic example. The problems grew in complexity such that they required the application of 16, 22, and 24 knowledge components, respectively. The example had the exact same underlying structure as the previous problem.

Predictions

The justification and step-focused prompts were designed to help augment an incomplete problem-solving schema. Meta-cognitive prompts were designed to facilitate the cognitive processes that repair a flawed mental model. If a student does not yet have a well-formed mental model, then the prompting may not appear to be effective to the student. Alternatively, most physics novices probably have an incomplete problem-solving schema. The justification and step-based prompts will probably be helpful to these students. Therefore, we predicted that students would attempt to use the prompts to study the examples for equal amounts of time on the first exposure. For the second example, we predicted that the justification and step-focused prompts would maintain a consistent level of engagement

with the examples, whereas the meta-cognitive group may not if they do not find their prompts helpful for learning.

What happens while studying examples should have an impact on problem-solving performance. If students in the justification and step-focused conditions learn more from the examples, then they should require less pedagogical assistance while solving problems. Specifically, students in the meta-cognitive condition are predicted to ask for more hints and bottom-out hints than the other two conditions.

Results

Studying Worked-out Examples

A mixed-model ANOVA was used to test the interaction of Condition (between-subjects factor: justification, meta-cognitive, step-focused) by Example (within-subjects factor: example 1 versus 2). One student was identified as an outlier ($z = 2.45$) and removed from this analysis. There was a main effect for Example ($F(1, 42) = 13.03, p < .05$) with participants spending more time studying the first example than the second, and the interaction between Example and Condition was reliable, $F(1, 42) = 5.00, p < .05$. To better understand the interaction, a simple-effects test was conducted. For the second example, both the step-focused ($F(1, 46) = 3.12, p = .08$) and justification conditions ($F(1, 46) = 3.10, p = .08$) demonstrated marginally longer study times than the meta-cognitive condition (see Example 2 of Figure 2).

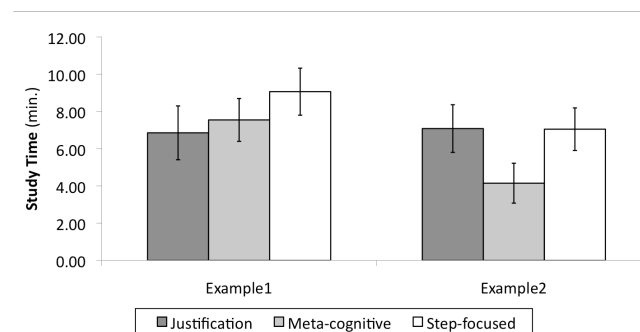


Figure 2. Mean (\pm SE) study time for each example

Collapsing across conditions, study times were negatively correlated with hint requests ($r(47) = -.37, p < .01$), and this was also true when the analyses were restricted to the study times for the second problem and the hint requests for the last problem, $r(47) = -.30, p < .05$. Moreover, this pattern of correlations was marginally significant for only the justification condition, $r(14) = -.46, p = .08$. This suggests that the more time spent on the examples, the fewer hint requests the students made while solving problems; therefore, study time was an indicator of learning from the examples.

Coached Problem Solving

The tradeoff between learning from examples and solving problems with tutorial support was tested next. It was predicted that the meta-cognitive condition would learn less from the examples than the other two conditions and thus rely more on the scaffolding supplied by tutoring system for successful problem solving. As expected, the meta-cognitive condition requested more hints than the justification-based condition, $F(1, 45) = 3.52, p = .07, d = .73$. Moreover, the difference was even more pronounced when just the bottom-out hints were examined (see Fig. 3). The meta-cognitive condition requested more bottom-out hints than the justification condition, $F(1, 45) = 5.49, p < .05, d = .75$.

A follow-up analysis revealed that the differences between the two conditions for bottom-out hint usage occurred during the second ($F(1, 45) = 5.52, p < .05, d = .72$) and third problems, $F(1, 45) = 4.52, p < .05, d = .74$.

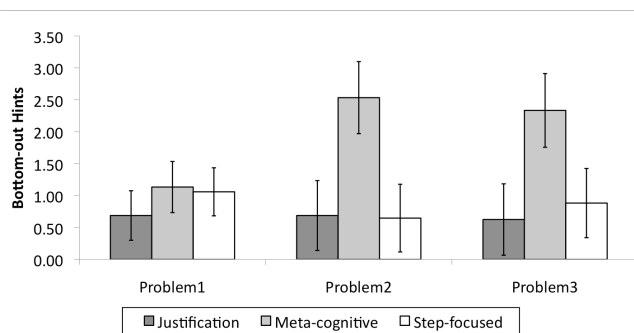


Figure 3. Mean (\pm SE) frequency of bottom-out hint requests

Learning Curves Analysis

The analyses reported above assume that learning did in fact occur. To support this assumption, we used an “embedded assessment” technique, which follows from theories of cognitive skill acquisition (Anderson, 1993). Specifically, as an individual becomes more proficient in deploying a skill, her error rate and amount of assistance decreases with each opportunity to apply that skill. In the present case, assistance came in the form of on-demand hints and immediate error flagging.

To evaluate learning, an *assistance score*, which is the sum of the number of hints and errors for each skill, was calculated for each opportunity (i.e., problem). The result is a learning curve that represents the assistance score as a function of opportunity. Note that a lower assistance score indicates more domain-relevant knowledge.

The knowledge component that we chose to focus on was the definition of the electric field, which is represented by writing the following equation: $F = qE$. We chose this particular knowledge component because it was used in all three problems, and it was identified as the first principle when solving the problems and examples.

The learning curve for the definition of the electric field can be found in Figure 4. There was a reliable, linear decrease in the amount of assistance needed to apply the

definition of the electric field, $F(1, 37) = 4.42, p < .05$. This suggests that learning did indeed occur during the experiment; however, this linear decrease was qualified by a linear-by-condition interaction, $F(2, 37) = 3.21, p < .05$. This suggests that the experimental conditions exhibited different learning rates. To further investigate this interaction, we analyzed the difference between conditions for each opportunity.

For the first opportunity, there was no difference between the three conditions, which suggests that the participants started the experiment with a similar level of background knowledge. For the second opportunity, the meta-cognitive condition demonstrated higher assistance scores than both the justification ($F(1, 37) = 12.38, p < .05$) and step-focused conditions, $F(1, 37) = 9.91, p < .05$. For the final opportunity, the meta-cognitive condition exhibited reliably higher assistance scores than the justification condition, $F(1, 37) = 7.09, p < .05$.

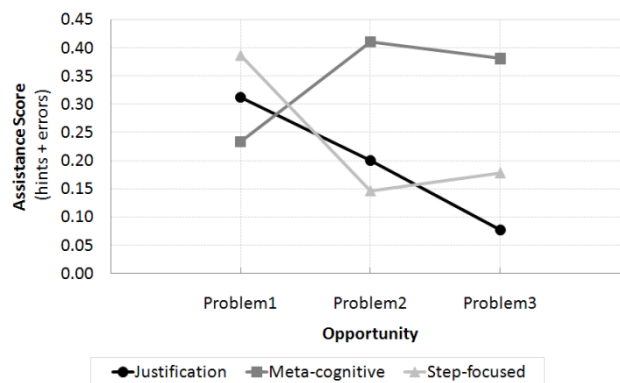


Figure 4. Learning Curves for the primary knowledge component ($F = qE$)

The learning-curve evidence suggests that the justification-based and step-focused prompts helped students learn from the examples better than the meta-cognitive prompting because their assistance scores decreased at a faster rate than the meta-cognitive condition.

Discussion

One of the goals of cognitive science is to develop theories that have a broad impact. The *self-explanation principle* has the potential to have a large effect on the design of learning environments and improving student learning in school settings. We designed and tested three different types of self-explanation prompts in the domain of solving electrodynamics problems with an intelligent tutoring system. We hypothesized that: 1) the students did not have many prior misconceptions for the target concepts (which was corroborated by our instructors’ intuition and a review of the physics education literature) and 2) that the most useful representation for solving problems with the Andes Physics Tutor is a problem-solving schema. The goal here was to examine how a cognitive-science principle translates

into an educational application, which critically depends on the interpretation of that principle (i.e., the theoretical assumptions upon which it is based) and the fit with the learning environment (i.e., learner factors).

The second goal of the project was a test of the Fit Hypothesis, which is the idea that there is an interaction between the student's prior knowledge, the cognitive processes that are evoked by different types of prompting, and the modifications those processes make to the domain representation.

The results provide strong support for the Fit Hypothesis showing that the justification and step-focused conditions benefited more from studying examples than the meta-cognitive condition. This result is consistent with the idea that the prompts in the justification and step-focused conditions facilitated the acquisition of a problem-solving schema and the generation of justifications for each step of the problem. The fact that both justification and step-focused performed equally well suggests that participants in the step-focused conditions spontaneously generated justifications for each step. In future work, we will test this hypothesis by examining the students' verbal protocols.

The fact that the meta-cognitive group relied more heavily on the tutoring support is consistent with the premise that students adopt (or ignore) instructional activities that they feel are helping them achieve their learning goals. For instance, the study times for the meta-cognitive condition dropped dramatically between the first ($M = 7.54$, $SD = 4.60$) and second ($M = 4.15$, $SD = 4.28$) example, $t(15) = 3.30$, $p < .05$, suggesting either that the students did not perceive the prompts as useful, or they did not have any existing prior knowledge to revise. Instead, they relied on learning from the on-demand hints. The justification and step-focused conditions requested fewer hints and bottom-out hints than the meta-cognitive prompting condition. The learning-curve analysis further supported these analyses.

There are two limitations to the current study. First, a closer examination of students' prior knowledge would strengthen the argument about the interaction between prior knowledge, representation, and cognitive processes. Future work should include such measures. Second, a full factorial study, crossing knowledge (prior vs. no prior knowledge) with prompting type (justification-based vs. meta-cognitive) would offer the strongest test of the Fit Hypothesis.

In conclusion, there is a great deal of importance and difficulty when interpreting learning principles and implementing them as instructional interventions. Research at the nexus of cognitive theory, technological tools, and classroom work is uniquely situated to address such multidisciplinary problems.

Acknowledgments

This work was supported by the Pittsburgh Science of Learning Center, which is funded by the National Science Foundation award number SBE-0354420.

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. (1993). *Rules of the mind*. Hillsdale, N.J.: L. Erlbaum Associates.
- 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. (1989). Learning from examples via self-explanations. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 251-282). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Chi, M. T. H., DeLeeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.
- 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.
- 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. In R. Luckin, K. R. Koedinger & J. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (Vol. 158, pp. 417-424). Amsterdam: IOS Press.
- Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, *69*(7), S12-S23.
- McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*(1), 1-30.
- Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(2), 427-440.
- VanLehn, K. (1998). Analogy events: How examples are used during problem solving. *Cognitive Science*, *22*(3), 347-388.
- 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.