# Explaining self-explaining: A contrast between content and generation

Robert G.M. HAUSMANN and Kurt VANLEHN
*Learning Research and Development Center, University of Pittsburgh*
*3939 O'Hara Street, Pittsburgh, PA, USA*

**Abstract.** Self-explaining has been repeatedly shown to result in positive learning outcomes for students in a wide variety of disciplines. However, there are two potential accounts for why self-explaining works. First, those who self-explain experience more content than those who do not. Second, there are differences in the activity of generating the explanations versus merely comprehending them. To compare these two accounts, the present *in vivo* classroom study, conducted in the PSLC physics LearnLab, attempted to contrast robust learning from generating explanations with the actively of studying instructional explanations. The students' learning activities (self-explaining vs. paraphrasing) were crossed with the completeness of the examples they studied (complete vs. incomplete). During a classroom period on electrodynamics, students alternated between solving problems and studying examples. On these problems, the self-explainers made fewer errors and asked for less help than paraphrasers. On homework problems done many days later, self-explainers displayed evidence of far transfer in a related, yet new domain (i.e., magnetism). In conclusion, prompting students to self-explain while studying examples, in an authentic classroom environment, can result in positive near- and long-term learning.

**Keywords**. self-explaining, paraphrasing, study strategies

## 1. Introduction

When students study instructional materials, including textbooks, worked-out examples, diagrams, and other multimedia materials, they may *self-explain* it, which means generating an explanation for what they have just read based on their knowledge or on content read earlier [1]. While self-explaining has consistently been shown to be effective in producing robust learning gains in several domains, including physics [2], the human circulatory system [3; 4], and algebra and geometry [5], the underlying sources of the effect are still being investigated. Identifying these sources is an important research objective because the design of effective educational practices and software can be guided by a deeper understanding of the process and outcomes of self-explaining. The present study attempts to add to this understanding, especially as it occurs in a real-world context: the classroom.

## 2. Explaining self-explaining

The first studies of self-explanation, which were based on analyses of verbal protocols, showed that the amount of self-explaining correlated strongly with performance on post-test measures of problem-solving performance [2; 6; 7]. Subsequent studies showed that students who were prompted to self-explain sentences in a scientific text learned more than students who were asked to paraphrase the sentences instead [2; 8; 9]. Other studies showed that self-explaining could be the elicited by computers [5; 9; 10; 11].

Because these studies compared self-explanation to the lack of any explanation at all, it is not clear why self-explanation produced learning. On the one hand, self-explaining generates additional information, namely the explanations themselves, that are not present in the instructional materials. Perhaps if students were simply given these explanations, they would learn just as much [12]. Alternatively, learning from self-explaining might arise from the activity of producing the explanations. Thus, if they were given the explanations, they would *not* learn just as much. In other words, is it merely *attending* to the explanations that matters, or is robust learning more likely to occur if students *generate* the explanations themselves? Let us label these hypotheses as follows:

1. *Attention*: learning from self-generated explanations should produce comparable learning gains as author-provided explanations, provided the learner pays attention to them. Both self-generated and author-provided explanations should exhibit better learning than no explanation.
2. *Generation*: learning from self-generated explanations should produce greater learning gains than author-provided explanations because they are produced from the students' own background knowledge; however, author-provided explanations should be comparable to no explanation.

There have only been a few empirical studies that attempt to separate the Attention hypothesis from the Generation hypothesis [13]. An exemplary case can be found in a study by Lovett [14] in the domain of permutation and combination problems. Lovett crossed the source of the solution (subject vs. experimenter) with the source of the explanation for the solution (subject vs. experimenter). For our purposes, only two of the experimental conditions matter: (1) the experimenter-subject condition, wherein the students self-explained an author's solution, and (2) the experimenter-experimenter condition, wherein the students self-explained an author's solution, whereas in the experimenter-experimenter condition, the students studied an author-provided explanation. Lovett found that the experimenter-experimenter condition demonstrated better performance, especially on far-transfer items. Lovett's interpretation was that the experimenter-experimenter condition was effective because it contained higher quality explanations than those generated by students. Consistent with this interpretation, when Lovett analyzed the protocol data, she found that the participants who generated the key inferences had the same learning gains as participants who read the corresponding inferences. Thus, of our two hypotheses, Lovett's experiment supports the Attention hypothesis: the content of self-explanations matters, while the source of the explanation does not.

Brown and Kane [15] found that children's explanations, generated 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 asked first, "Why would a furry caterpillar want to look like a snake?" and if that did not 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. This result is consistent with the Generation hypothesis, which is that an explanation is effective when the student generates it. However, the students who were told the rule may not have paid much attention to it, according to Brown and Kane.

In summary, one study's results are consistent with the Attention hypothesis, and the other study's results are consistent with the Generation hypothesis, but both studies confounded two variables. In the Lovett study, the student-produced and author-provided explanations were of different qualities. In the Brown and Kane study, the students in the author-provided explanations condition may not have paid much attention to the explanations.

To contrast the Generation and Attention hypotheses, we conducted an experiment in the physics LearnLab (www.learnlab.org), which is a course that is designed to conduct rigorous, *in vivo* experiments on issues related to robust learning. Robust learning is defined in three parts. First, learning is considered robust when the knowledge is retained over a significant period of time. Second, robust learning is the opposite of inert knowledge in the sense that students are able to broadly apply their knowledge to other problems within the same class. Finally, robust learning can be applied to a different, yet related, domain. Robust learning is contrasted with "normal" learning, which is the performance on problems that are near transfer and retained over a relatively short duration.

## 3. Materials

The training materials[1] used for the present experiment were developed in association with one of the LearnLab instructors and two other physicists. The experiment covered the domain of electrodynamics, with an emphasis on the forces acting on a charged particle due to the presence of an electric field. The homework problems were near- and far-transfer versions of the training problems, along with problems from the magnetism chapter. Magnetism was used as a test-bed for studying transfer across domains. The data were collected in the LearnLab, as opposed to the laboratory, because the realism of the classroom increases the generalizability of the results without sacrificing randomization or control over extraneous variables.

### *3.1. Design*

The experiment was a 2 x 2 between-subjects design, which crossed two independent variables: Study Strategy (paraphrasing vs. self-explaining) and Example Type

[1] For a copy of the experimental materials, visit: http://andes3.lrdc.pitt.edu/~bob/mat/exper1.html

(complete vs. incomplete). When students studied a complete example, they were given an author-provided explanation, and both the paraphrasing (PP) and self-explanation (SE) instructions were intended to insure that they paid attention to it. When students studied an incomplete example, the self-explanation prompts were intended to elicit self-explanation, while the paraphrase instructions were intended to block it. The instructions were modeled after earlier studies [2; 9], where verbal protocols confirmed that they had the intended effects of eliciting or suppressing self-explanations. Students were block randomized into one of the following 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 two hypotheses make the following predictions:

- *Attention*: SE-complete = SE-incomplete = PP-complete > PP-incomplete
- *Generation*: SE-complete = SE-incomplete > PP-complete = PP-incomplete

## 4. Procedure

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. The experiment took place in one of the open class periods, which were approximately 110 minutes in duration.

The students were introduced to the experiment and shown the instructions. Then students were prompted to solve the first problem, which was a warm-up problem to get the students acquainted with the Andes interface. Andes (www.andes.pitt.edu) is an intelligent tutoring system, created for helping students solve homework problems from the first two semesters of physics [16]. After solving the first problem, the students then studied the first example. This process, alternating between solving problems and studying examples [17], 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 paraphrase or self-explain at the end of each segment. To capture their verbalizations, each student wore 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 screen-logging 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 text-only log file of each action taken in the Andes interface. In addition to the in-class experimental session, log files from the assigned Andes homework problems were made available to the researchers. Problems from the electric and magnetic fields chapters were of particular relevance. The course instructor also provided the paper-and-pencil chapter exam that included a problem with a charged particle moving in an electric field.

## 5. Results

This section is broken down into three sub-sections, corresponding to the type of learning that was measured: normal and robust, which included within- and between-domain, far-transfer items. Normal learning was defined as problem-solving performance on the items administered on the day of the experiment. Robust within-domain, far-transfer learning took place separately from the experiment, while the students solved either homework (Andes) or exam (paper-and-pencil) problems. Robust between-domain, far-transfer learning was performance on Andes homework problems, taken from a topic that was separate from the experimental materials (i.e., magnetism).

Because Andes insured that students always found a correct solution to the problem, our main dependent measure was their normalized assistance score on the problem, which was defined as the sum of all the errors and requests for help on that problem divided by the number of entries made in solving that problem. Thus, lower assistance scores indicate that the student derived a solution while making fewer mistakes and getting less help, and thus demonstrating better performance and understanding.

### *5.1. Normal learning: Problems solved during the experimental session*

To ensure equivalent experimental conditions, we analyzed performance on the warm-up problem, which revealed no reliable differences, $F(3, 99) = .37$, $p = .78$. In terms of normal learning, normalized assistance scores were contrasted by averaging over individuals for all three problems in the training set. Using a repeated-measures Analysis of Variance (ANOVA), a main effect for Study Strategy was observed, with the self-explanation condition demonstrating lower normalized assistance scores than the paraphrase condition (see Figure 1), $F(1, 73) = 6.19$, $p = .02$, $\eta_p^2 = .08$. This result is consistent with the Generation hypothesis.

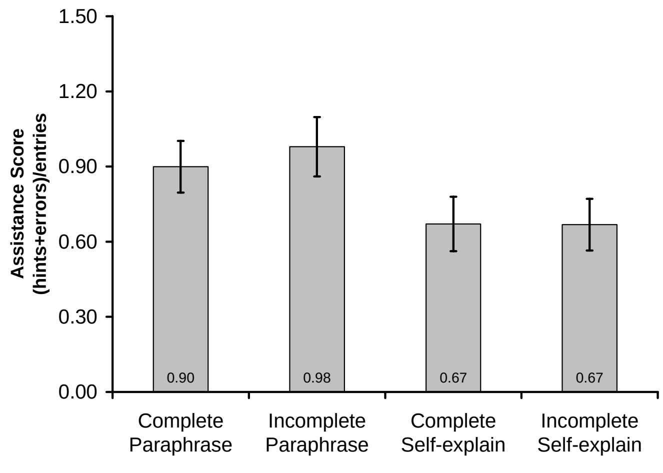


**Figure 1.** Means and standard errors on the three training problems.

### *5.2. Robust learning: Within-domain, far transfer*

On the chapter exam, administered 29 days after the experiment, there were neither reliable main effects nor an interaction; however, post-hoc comparisons using the

Fischer LSD test revealed that the complete self-explanation condition (*M* = 90.83, *SD* = 9.96) had a marginally higher score than the complete paraphrase condition (*M* = 73.00, *SD* = 22.51) (LSD, $p = .06$). The reason why a difference was not observed could be due to the insensitivity of the measure or that the students' preparation for the exam washed out any potential differences between conditions.

To overcome these potential limitations, we analyzed the students' performance on a homework problem that was isomorphic to the chapter exam, in the sense that they shared an identical deep structure (i.e., both analyzed the motion of a charged particle moving in two dimensions). The homework problem was solved after the training but before the chapter exam. The marginal effect observed on the chapter exam was replicated on the isomorphic homework problem. There was a reliable main effect for Study Strategy, with the self-explanation condition demonstrating lower normalized assistance scores than the paraphrase condition, $F(1, 27) = 4.07$, $p = .05$, $\eta_p^2 = .13$ (see Figure 2). This result is also consistent with the Generation hypothesis.

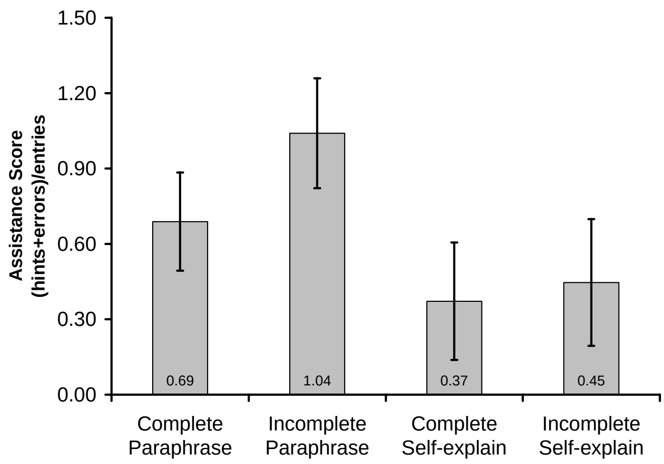


**Figure 2.** Means and standard errors on a homework electric-field problem.

*5.3. Robust learning: Between-domain, far transfer*

Another component of robust learning is to examine performance on problems in a different domain. To measure far transfer between domains, we analyzed homework performance in the magnetism chapter. We chose a magnetism problem that was somewhat similar to the electrodynamics training problems, although there remained major differences due to the differences in the task domain. The similarities were that both the principle knowledge component from electrodynamics (i.e., the definition of an electric field: $\mathbf{F_e} = q\mathbf{E}$) and magnetism (i.e., the vector expression for a magnetic force on a charged particle moving in a magnetic field: $\mathbf{F_b} = q\mathbf{v}\times\mathbf{B}$) were vector equations. Second, both problems included an opposing force that was equal in magnitude, but opposite in direction; thus, there was no acceleration in either case.

Using an ANOVA, a main effect for Study Strategy was observed, with the self-explanation condition demonstrating lower normalized assistance scores than the paraphrase condition (see Figure 3), $F(1, 46) = 5.22$, $p = .03$, $\eta_p^2 = .10$. Once again, the results favor the Generation hypothesis.

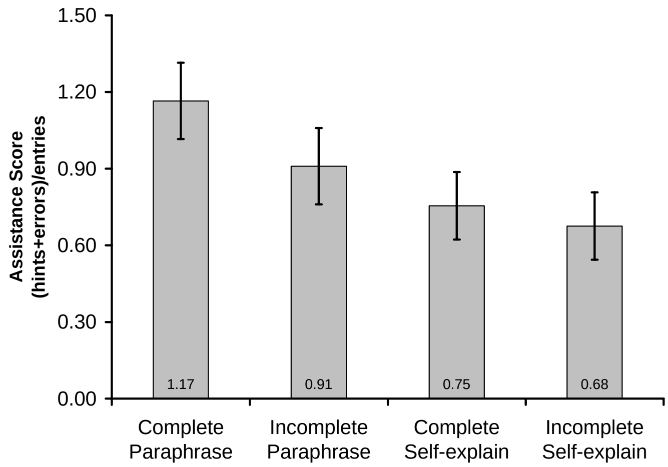


**Figure 3.** Means and standard errors on a homework magnetic-field problem.

## 6. Conclusions

Prior to this study, it appeared that the self-explanation effect could be due simply to attending to the content of the explanations, and that it did not matter whether the explanations were provided by an author or produced by the student. This hypothesis, which we called the Attention hypothesis, was not supported by our data. When students paraphrased high-quality explanations, we found less learning than students who were prompted to generate their own explanations of examples which omitted the author-provided explanations. Paraphrasing was used so that we could be reasonably sure that they at least paid some attention to the explanations. Thus, the evidence favors the Generation hypothesis, which asserts that the important variable for learning was the *process* of producing an explanation. In future work, we hope to gain insights into the underlying process by analyzing the recordings collected during the study.

This study also demonstrates that a fairly robust finding from the laboratory [2] could be imported into the classroom. The evidence suggests that prompting for self-explaining can be replicated in the real world, even when the conditions are such that the students are prompted to self-explain worked-out video examples in a noisy, complex environment such as a classroom. Prompting for self-explaining had a positive influence on normal learning. Normal learning consisted of a short retention interval, and problems that were designed to be isomorphic to the worked-out examples. The positive influence of self-explaining on learning has been shown on normal learning in previous laboratory studies [4]; however, the present study takes this result one step further by evaluating the effect on robust learning measures, including features such as long retention intervals, far transfer within and between domains.

## 7. Acknowledgements

This project was funded by the Pittsburgh Science of Learning Center under grant 0354420 from the NSF. The authors would like to thank the anonymous reviewers for their helpful comments; Robert Shelby and Brett van de Sande for their assistance in material development; Michael Ringenberg, Min Chi, and Anders Weinstein for their technical expertise; and we are extremely grateful to Donald Treacy, John Fontanella, and Mary Wintersgill for graciously allowing us to collect data in their classrooms.

## 8. References


[1] 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.

[2] 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.

[3] Ainsworth, S., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. Cognitive Science, 27(4), 669-681.

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

[5] 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.

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

[7] Bielaczyc, K., Pirolli, P., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. Cognition and Instruction, 13(2), 221-252.

[8] Magliano, J. P., Wiemer-Hastings, K., Millis, K. K., Munoz, B. D., & McNamara, D. S. (2002). Using latent semantic analysis to assess reader strategies. Behavior Research Methods, Instruments & Computers, 34(2), 181-188.

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

[10] 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.

[11] 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.

[12] Cavalli-Sforza, V. (1999). Constructed vs. received representations for scientific argument: Implications for learning and coaching. In M. Hahn & S. C. Stoness (Eds.), Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society. Mahwah, NJ: Erlbaum.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.