Is Self-Explanation Always Better? The Effects of Adding Self-Explanation Prompts to an English Grammar Tutor

Ruth Wylie (rwylie@cs.cmu.edu)
Human-Computer Interaction Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15217 USA

Kenneth R. Koedinger (koedinger@cmu.edu)
Human-Computer Interaction Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15217 USA

Teruko Mitamura (teruko@cs.cmu.edu)
Language Technologies Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15217 USA

Abstract

Several studies have demonstrated the benefits of self-explanation on learning well-defined domains like math, biology, and physics. However, these findings have yet to be replicated in probabilistic domains like second language acquisition. Working with adult English as a Second Language students (n=61) within the domain of the English article system (i.e. teaching students the difference between *a* dog vs. *the* dog) we conduct the first experimental study of the effects of prompting self-explanation on second language grammar acquisition. We compare two different modes of self-explanations (free-response and menu-based), each implemented in an intelligent tutoring system, to a control tutor with no explicit self-explanation prompts. Students in all conditions show significant learning gains but contrary to theoretical predictions, the self-explanation tutors did not lead to better learning over the no self-explanation condition. We discuss why and under what specific conditions target-specific practice without self-explanation may be a more effective instructional strategy.

Keywords: Self-Explanation Effect; Computer Assisted Language Learning; ESL Grammar Learning

Introduction

Self-explanation has been shown to be a successful learning strategy for multiple domains, contexts, and learners. One limitation of the existing work is the domains in which it has been tested have all been math and science domains like biology (Chi, et al., 1994), physics (Chi, 1989; Conati & VanLehn, 2000), and geometry (Aleven & Koedinger, 2002), and, to the best of our knowledge, there have never been any experimental studies on the effects of self-explanation on second language grammar acquisition. Thus, an open question exists: is self-explanation truly domain independent (Roy & Chi, 2005) or are there constraints to its applicability?

In the original self-explanation studies, Chi et al. (1989) examined students' spontaneous self-explanations of a physics text. This work revealed a positive correlation between the number and type of self-explanations and student learning. In subsequent experimental studies, Chi et al. (1994) showed that students who were prompted to self-explain demonstrated greater learning gains than those who were not. Furthermore, Aleven and Koedinger (2002) demonstrated that prompting self-explanations can be an effective learning strategy even when students only select a general problem-solving principle. Within the second language acquisition community, there is a large body of research that looks at implicit versus explicit instruction. A meta-analysis of the relative effectiveness of different types of second language instruction revealed that treatments involving explicit focus on rules were more effective than those that did not (Norris & Ortega, 2000). Thus, self-explanations, which highlight explicit rules, may be beneficial for the second language learner.

Our goal was to see if the success of self-explanation could be replicated within second language acquisition. To this end, we developed two tutoring systems with different types of self-explanation prompts and compared student learning gains and learning efficiency scores to a control tutor that had no explicit self-explanation prompts. Results show that while students in all three conditions demonstrate significant pre-post learning gains, students in the self-explanation conditions did no better than those in the control group. In fact, a significant learning efficiency by tutor condition interaction reveals that there may be limits to the benefits of self-explanation.

Adding Self-Explanation to an Existing Tutor

Self-explanation prompts were added to an existing tutoring system designed to teach the English article system (teaching students the difference between "*a* dog" and "*the* dog"). In the existing system (Figure 1), developed using the Cognitive Tutoring Authoring Tools (Koedinger, et al., 2004), students select an article (*a*, *an*, *the*, or *no article*) from a drop-down menu to complete the sentence. They receive immediate feedback on their selections (the answer

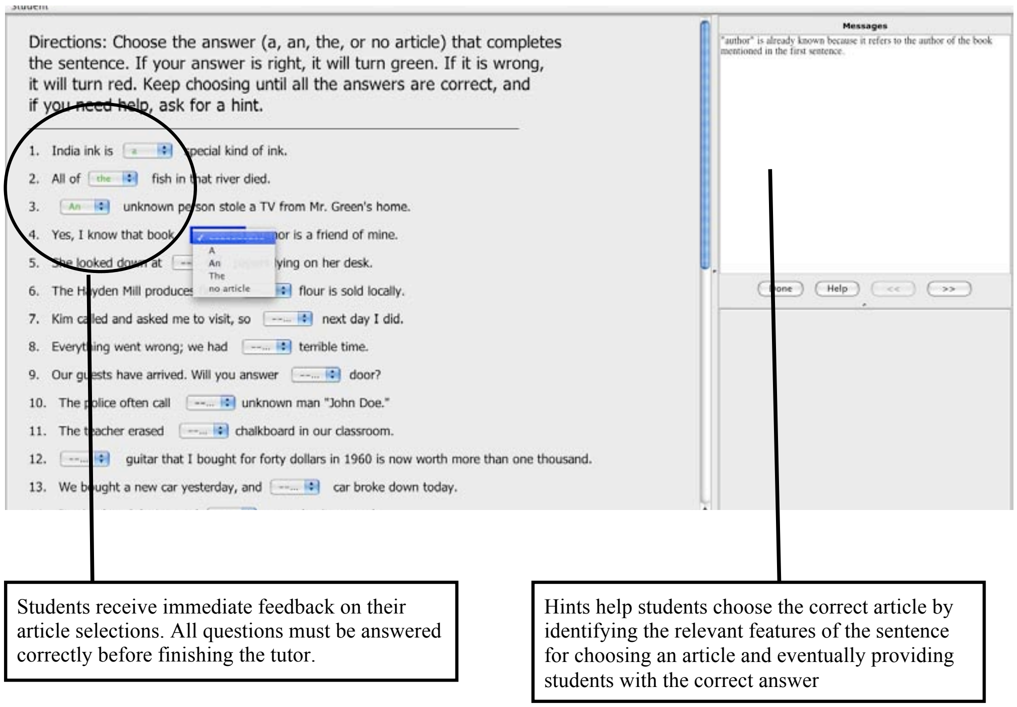

Figure 1: No self-explanation tutor. Students select an article to complete the sentence but are not prompted to self-explain.

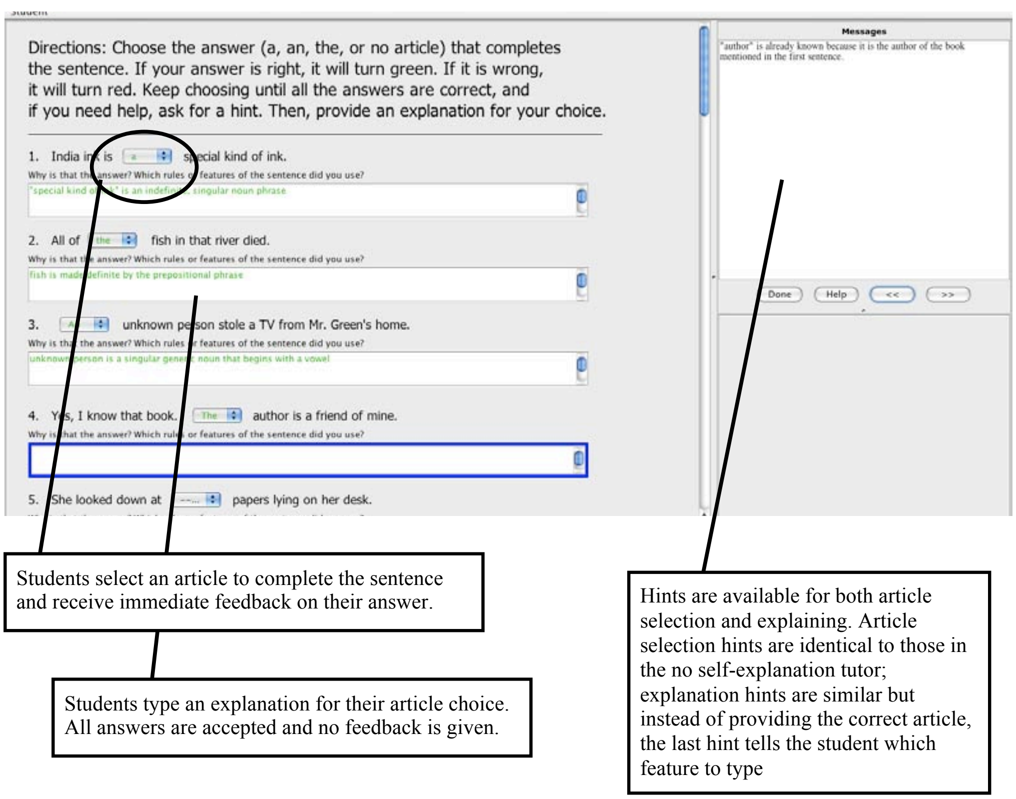

Figure 2: Free response self-explanation tutor. Students select an article to complete the sentence and then provide a written explanation for the answer.

turns green if it is correct, red if it is incorrect) and have access to a series of on-demand hints. The hints first identify the relevant features of the sentence and eventually provide students with the correct answer (Table 1). This tutor served as the instruction for the no self-explanation (control) condition of the study. To investigate the effects of self-explanation, we enhanced this tutor by adding two different modes of explaining to create a free-response self-explanation tutor and a menu-based self-explanation tutor.

Table 1: Example hint sequence provided for students making an article selection for the sentence:
Yesterday, I bought a TV. Today, ___ TV broke.

Hint text for **article selection** tasks	
1	TV has already been mentioned.
2	When a noun has already been mentioned, use "the".
3	Please select "the" from the highlighted menu.

Free response self-explanation tutor

The prompts for the free response self-explanation tutor were based on those used in the Chi et al. biology study (1994). In that study, students were prompted to verbally explain what they had just read (the text was presented one sentence at a time) and were not constrained in the length or content of their explanations. Following this approach, in the free response self-explanation tutor, students were asked: "Why is that the answer? Which rules or features did you use to make your choice?" Students type their responses in a textboxes. All answers are accepted, and no feedback on their explanation is given (Figure 2). Students have access to hints to aid with the self-explanation step (Table 2). The hints, similar to the hints for article selection, identify the relevant features of the sentence and then provide the rule that dictates which article should be used.

Table 2: Example hint sequence provided for students explaining the sentence:
*Yesterday, I bought a TV. Today, **_the_** TV broke.*

Hint text for **explanation selection** tasks	
1	"TV" was mentioned in the first sentence.
2	Since "TV" was already mentioned, it is definite.
3	Please enter "The noun has already been mentioned" from the highlighted menu.

Menu self-explanation tutor

One of the potential disadvantages of the free response method of self-explanation is that we cannot easily provide feedback to students on their explanations. However, if students were to select a rule or explanation from a given list, as they did in the self-explanation supported Geometry Cognitive Tutor (Aleven & Koedginer, 2002), the tutor could provide relevant feedback and insure their explanation is correct before continuing. In the Geometry Cognitive Tutor, students explained their steps by choosing the relevant rule from the provided glossary. In a similar fashion, students using the menu-based article tutor choose an explanation for their article choice from a drop-down menu (Figure 3). Students receive immediate feedback and again, identical to the free response self-explanation tutor, have access to hints.

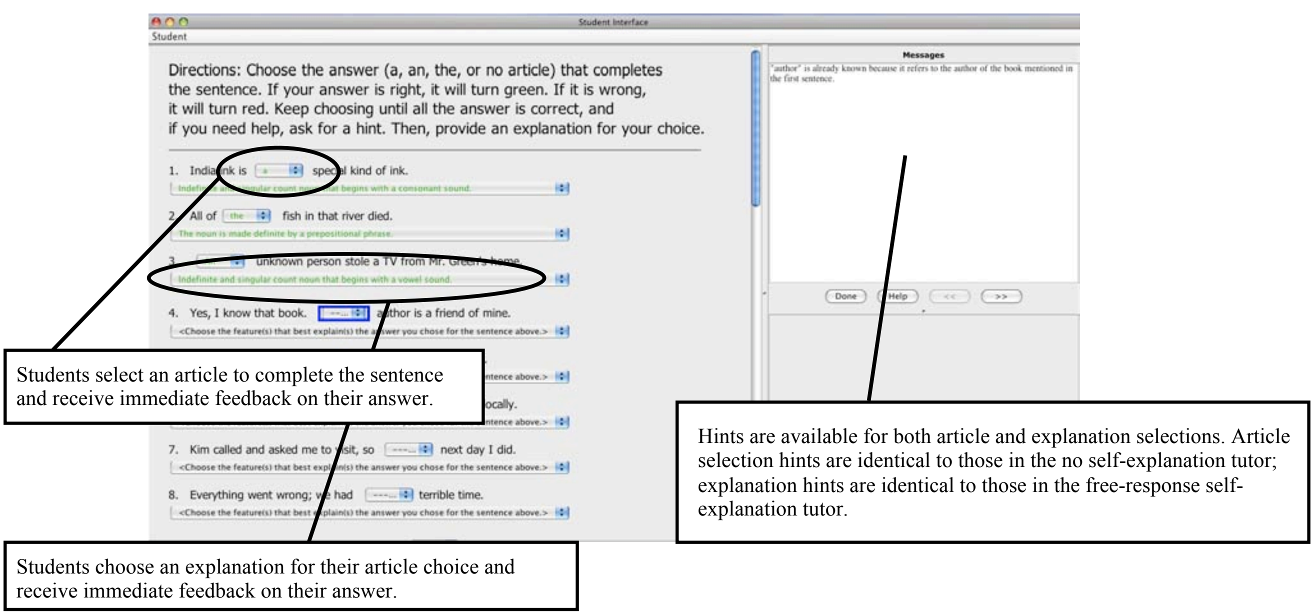

Figure 3: Menu self-explanation tutor. Students select an article to complete the sentence and the rule/feature that best explains their choice.

	Pretest (n=61)			Posttest (n=61)		
Tutor Condition	Article Tasks (16 items)	Explanation Tasks (8 items)	*Total (24 items)*	Article Tasks (16 items)	Explanation Tasks (8 items)	*Total (24 items)*
Free Response	58.9% (14.3)	37.5% (27.7)	*52.0% (15.5)*	73.5% (11.8)	42.9% (30.3)	*63.3% (14.6)*
Menu	69.7% (18.7)	48.7% (23.9)	*62.7% (17.4)*	76.6% (15.0)	50.0% (28.0)	*67.8% (15.9)*
No Self-Explanation	60.4% (13.4)	39.3% (19.9)	*53.4% (11.3)*	74.4% (11.5)	41.1% (26.9)	*63.3% (14.2)*
Total	**62.8% (16.0)**	**41.6% (24.1)**	***55.8% (15.3)***	**74.8% (12.7)**	**44.5% (28.2)**	***64.7% (14.8)***

Table 3: Mean and standard deviations for posttest scores by assessment category and tutoring condition. Students in all conditions showed significant pre to posttest gains for article items only.

Methodology

Participants were adult students enrolled in one of three levels (intermediate, high-intermediate, advanced) of an English as a Second Language (ESL) grammar course. Genders were equally represented and the students came from a variety of first languages. Students began with a wide range of initial competency. Pretest scores ranged from 25% to 100.0% (M = 57.1%, SD = 16.7, n = 63). Out of a total of 68 participants, 5 chose not to have their data collected, and 2 scored greater than 90% on the pretest and thus were removed from analysis leaving us with 61 participants, 21 in the free-form and no self-explanation conditions and 19 in the menu self-explanation condition).

The study was conducted within the University of Pittsburgh's English Language Institute. Students were enrolled in ESL grammar courses and participated in the study as part of their regular coursework. Students in the intermediate (n=15) and high-intermediate (n=42) courses completed the tutor and assessments as an in-class activity, while students in the advanced course (n=4) completed them as a homework assignment. All students completed a computer-based pre and posttest that consisted of article-only and article with explanation items. In the article-only items, students chose an article from a dropdown menu to complete the sentence. In the article with explanation items, students chose an article to complete the sentence and then chose the feature or rule that explained their answer. No hints were available during the tests, and students did not receive feedback on their answers. Students were randomly assigned to tutor condition. In an attempt keep time on task about equal, students in the no self-explanation condition completed three times as many article selection tasks (84 sentences vs. 28 sentences in the self-explanation tutors). The decision to have students complete more sentences was made after pilot data showed that completing 28 no self-explanation items took about one third the time as completing 28 matched self-explanation items. We chose to control for time on task versus number of items in order to increase ecological validity. Our intervention was designed and carried out during a regular class period and thus it was important for the duration of the intervention to approximate the duration of class. Furthermore, in previous self-explanation studies (Aleven and Koedinger, 2002), controlling for number of items rather than time on task lead to challenges in interpreting the results. Since self-explanation requires additional time, had we chosen to control for number of items, any observed effects of self-explanation would be confounded with an increase amount of time spent using the tutor.

Results

What are students learning with the article tutors?

The assessment items were divided into two categories: target items (the article selection tasks) and explanatory items (the explanation selection tasks) (Table 3). As the goal of the tutoring unit was to increase performance on the target tasks, we were less concerned with how students performed on the explanatory tasks. In fact, native speakers usually can't explain these rules but have no trouble using articles. A repeated measures ANOVA with score on target items as the dependent variable reveals a significant main effect for test time ($F(1, 58) = 42.6$, $p < 0.001$) indicating a significant pre to posttest gain (Figure 4)[1].

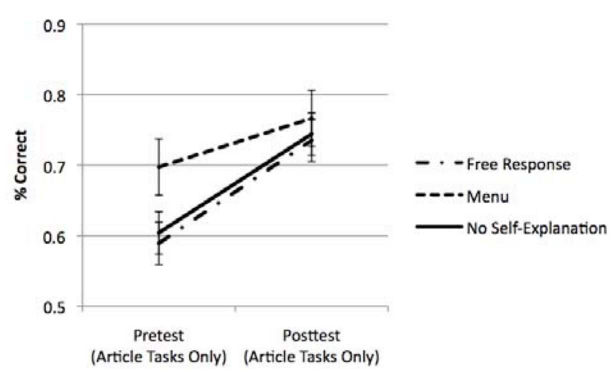

Figure 4: Students demonstrated significant pre to posttest gain for article selections with a marginally significant effect of condition.

[1] All analyses were repeated including all participants (n=63). The results revealed a similar pattern to those presented; and there was no difference in the interpretation of results.

A similar analysis for the explanation tasks shows no increase from pre to posttest regardless of condition ($F(1,58) = 1.27$, $p = 0.27$). While not surprising for the no self-explanation group since they did not receive practice explaining or the free response self-explanation group since they did not receive feedback on their explanations, it is surprising that even students in the menu tutoring condition, who had practice selecting explanations and immediate feedback on their choices, did not improve in their ability to select the correct rule that explained their answer.

How much time did students spend using the tutors?

When evaluating classroom interventions, another important factor is the amount of time it takes students to complete the instruction. While the instruction was designed to keep time-on-task close, there was a marginal difference between the conditions in the amount of time it took students to complete the tutors ($F(2, 58) = 2.90$, $p = 0.063$). Post-hoc Tukey HSD tests revealed that students using the menu tutor completed the instruction the fastest (M = 13.3 minutes, SD = 6.0) but not significantly faster than those who used the no self-explanation tutor ($p = 0.682$, M = 15.2 minutes, SD = 6.8), and the no self-explanation tutor was not significantly faster than the free-response tutor ($p = 0.270$, M = 18.5, SD = 7.8) However, the menu self-explanation tutor was completed marginally faster than the free-response tutor ($p = 0.056$) (Table 4).

Table 4: Pairwise comparison of time spent using the tutor by condition.

Condition (i)	Condition (j)	Mean Difference (i-j)	p-value
No Self-Exp	Free Response	-3.35	0.270
	Menu	1.84	0.682
Menu	Free- Rsponse	-5.19	0.056

Another metric used to compare the effectiveness of the tutoring conditions is learning efficiency. Efficiency scores[2] combine time-on-task and learning gains into a single measure. In order to account for varying pretest scores, normalized gain scores[3] were used. Multiple linear regression with efficiency score as the dependent variable, condition as the independent variable and pretest as a covariate reveals a significant pretest by condition interaction ($F(2, 60) = 3.54$, $p = 0.036$) and a marginally significant main effect of condition ($F(2, 60) = 2.49$, $p =0.092$). As the scatterplot shows (Figure 5), students with high pretest scores tended to benefit more from the no self-explanation tutor; whereas students with lower pretest scores tended to be more efficient while using the free-response tutor.

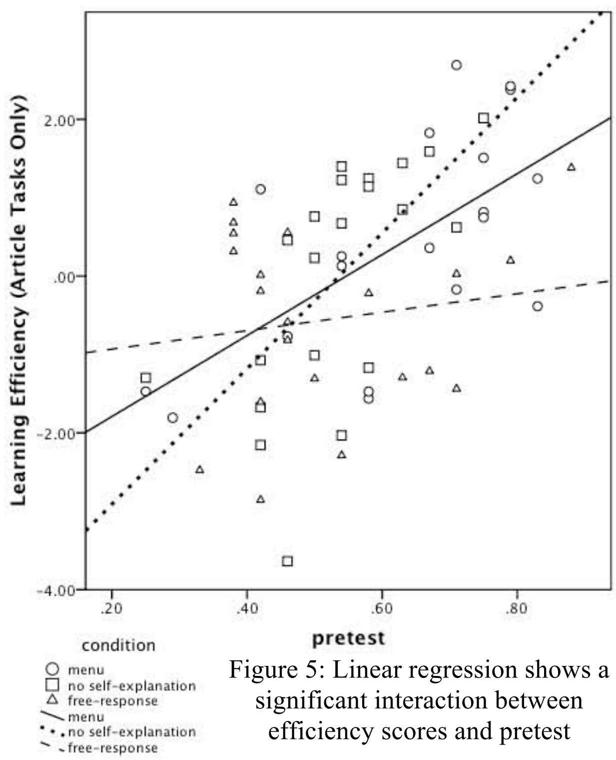

Figure 5: Linear regression shows a significant interaction between efficiency scores and pretest

Discussion

The current results suggest that there are limitations to the benefits of self-explanation. One reason could be that the act of generating and selecting explanations added extraneous cognitive load to the task. The no self-explanation condition simply provided students with concentrated practice of the target items, thereby reducing extraneous load. In addition to taking time away from practice on the target items, the act of self-explaining might have actually hindered noticing all the relevant features for choosing the correct article. Research on verbal overshadowing might help explain this claim. Verbal overshadowing is the effect that those who describe a previously seen face do worse on identifying that face than those in a no-description control condition. One hypothesis is that people who don't provide a description approach the identification task in a global manner while those who generate or read a description narrow their focus to specific features (Meissner & Brigham, 2000). If those features prove to be unreliable cues, performance declines. Similarly, when selecting articles, the act of generating or selecting an explanation may cause students to ignore less salient, but important, cues and make incorrect article decisions.

However, it is also possible that it is the inherent differences between the domains (or between the particular knowledge goals within the domains) that are driving the results. Second learning language is different from learning the math or science principles (as opposed to facts or notations) that where the target of past self-explanation

[2] Efficiency = Zscore(gain) – Zscore(time)

[3] For positive gains, normalized gain = (posttest-pretest)/(1-pretest), for negative gains, normalized gain = (posttest-pretest)/pretest.

studies. In his review article, DeKeyser (2005) notes that learning ESL articles is difficult because they are abstract and novel. Articles are abstract in the sense that learners have a difficult time understanding the meaning of the article and novel when the student's first language does not have articles or has a very different article system. Theoretically, a successful instructional intervention would be one that explicitly addresses these sources of difficulty. Perhaps the reason why the self-explanation tutors were not as beneficial is because the explanations highlighted key features of the sentence but did little to address the meaning of the article itself or how the article affects the meaning of the sentence.

It appears that for procedures that are difficult to explain (i.e., those for determining which article to use), receiving more practice opportunities with less reflective instructional practice (i.e., 3 times as many items in the no self-explanation condition) is better than fewer opportunities but more reflection per item. Prior self-explanation studies involved more complex procedures that can be explained with well-defined principles that are articulated in math and science textbooks. For these complex, principle-based procedures, using fewer items with more reflection appears to yield more effective and equally efficient learning.

The significant aptitude-treatment interaction (shown in Figure 5) indicates that even for article knowledge, some level of example study and reflection may be useful for early learners. Until such learners have a reasonably high chance of getting practice items correct, mere practice may be inefficient and some early reflective example study may be in order (cf., Koedinger, Pavlik, McLaren & Aleven, 2008).

This work highlights the need to continue investigating the self-explanation effect in new and different domains. It suggests there may be limitations to its applicability. Additionally, it is important to understand the source of difficulty within a domain and identify how self-explanations may or may not address it. More generally, it indicates that potential general principles of learning and instruction may only be effective in combination with a detailed cognitive task analysis of the domain knowledge and awareness of relevant boundary conditions. More research is needed to further specify those boundary conditions and relate them to basic understanding of cognitive processes.

Acknowledgements

Special thanks Jim Rankin, Jonathan Sewell, Erin Walker, and our partner teachers for their invaluable help. Also thanks to our anonymous reviewers for their helpful suggestions and comments. This work was supported in part by the Pittsburgh Science of Learning Center which is funded by the National Science Foundation award number SBE-0354420 and by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B040063 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Aleven, V. & Koedinger K. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, *26*, 147-179.

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

Conati, C., & VanLehn, K. (2000). Further results from the evaluation of an Intelligent Computer Tutor to Coach Self-Explanation. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 304-313.

DeKeyser, R. (2005) What Makes Learning Second Language Grammar Difficult? A Review of Issues. *Language Learning*, 55:s1, 1-25.

Koedinger, K., Aleven, V., Heffernan. T., McLaren, B. & Hockenberry, M. (2004). Opening the Door to Non-Programmers: Authoring Intelligent Tutor Behavior by Demonstration. *Proceedings of 7th Annual Intelligent Tutoring Systems Conference*. Maceio, Brazil.

Koedinger, K. R., Pavlik Jr., P. I., McLaren, B. M., & Aleven, V. (2008). Is it better to give than to receive? The assistance dilemma as a fundamental unsolved problem in the cognitive science of learning and instruction. In B.C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, 2155-2160.

Meissner, C., & Brigham, J. (2000) A Meta-analysis of the Verbal Overshadowing Effect in Face Identification. *Applied Cognitive Psychology*, *15*, 603-616.

Norris, J., & Ortega, L. (2000) Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 157-213.

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

VanLehn, K., & Jones, R. (1993). What mediates the self-explanation effect? Knowledge gaps, schemas or analogies? In M. Polson (Ed.), , 1034-1039. Hillsdale, NJ: Erlbaum.