Examining the Generality of Self-Explanation

Ruth Wylie

September 2011

CMU-HCII-11-105

Human-Computer Interaction Institute
School of Computer Science
Program for Interdisciplinary Education Research
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Thesis Committee

Kenneth Koedinger, Carnegie Mellon University (Co-chair)
Teruko Mitamura, Carnegie Mellon University (Co-chair)
Sharon Carver, Carnegie Mellon University
Albert Corbett, Carnegie Mellon University
Carolyn Rosé, Carnegie Mellon University

Submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

This research was supported in part by the Pittsburgh Science of Learning Center, funded by National Science Foundation award number SBE-0354420 and by the Program for Interdisciplinary Education Research, funded by the Institute of Education Sciences grant R305B040063.

Keywords

Prompted self-explanation, tutored practice, worked examples, analogical comparisons, second language acquisition, English article system, educational technology, computer-assisted language learning

Abstract

Prompting students to self-explain during problem solving has proven to be an effective instructional strategy across many domains. However, despite being called "domain general", very little work has been done in areas outside of math and science. In this dissertation, I investigate whether the self-explanation effect holds when applied in an inherently different type of domain, second language grammar learning.

Through a series of *in vivo* experiments, I tested the effects of using prompted self-explanation to help adult English language learners acquire the English article system (e.g., teaching students the difference between "I saw *a* dog" versus "I was *the* dog"). In the pilot study, I explored different modalities of self-explanation (free-form versus menu-based), and in Study 1, I looked at transfer effects between practice and self-explanation. In the studies that followed, I added an additional deep processing manipulation (Study 2: analogical comparisons) and a strategy designed to increase the rate of practice and information processing (Study 3: worked example study). Finally, in Study 4, I built and evaluated an adaptive self-explanation tutor that prompted students to self-explain only when estimates of prior knowledge were low. Across all studies, results show that self-explanation is an effective instructional strategy in that it leads to significant pre to posttest learning gains, but it is inefficient compared to tutored practice. In addition to learning gains, I compared learning process data and found that both self-explanation and practice lead to similar patterns of learning and there was no evidence in support of individual differences.

This work makes contributions to learning sciences, second language acquisition (SLA), and tutoring system communities. It contributes to learning sciences by demonstrating boundary conditions of the self-explanation effect and cautioning against broad generalizations for instructional strategies, suggesting instead that strategies should be aligned to target knowledge. This work contributes to second language acquisition theory by demonstrating the effectiveness of computer-based tutoring systems for second language grammar learning and providing data that supports the benefits of explicit instruction. Furthermore, this work demonstrates the relative effectiveness of a broad spectrum of explicit learning conditions. Finally, this work makes

contributions to tutoring systems research by demonstrating a process for data-driven and experiment-driven tutor design that has lead to significant learning gains and consistent adoption in real classrooms.

Acknowledgements

Through the dissertation process, I have been incredibly fortunate to be surrounded by amazing teachers, friends and collaborators.

To my advisors, Ken Koedinger and Teruko Mitamura, thank you for the guidance, patience and expertise. You're brilliant mentors and your commitment to education and research is inspiring. This work has also benefitted greatly through the contributions of my committee members, Sharon Carver, Albert Corbett, and Carolyn Rosé. Thank you for your insightful comments and questions.

In addition to my advisors and committee members, I was also fortunate to have the help and support of Vincent Aleven, Howard Seltman, Anind Dey, Jennifer Mankoff, David Klahr, and Chuck Perfetti.

Jo Bodnar, Queenie Kravitz, and Audrey Russo, thank you for helping with all the logistics and for doing so with a smile. Jonathan Sewall, Chang-Hsin Chang, Martin van Velsen, Octav Popescu, Kyle Cunningham, Jim Rankin, Alida Skogsholm, and Brett Leber helped tremendously with tutor development and DataShop support. Gail Kusbit and Michael Bett, thanks for keeping the PSLC running. Dorolyn Smith and Alan Juffs, thank you for opening up your classrooms and your invaluable pedagogical insight. Melissa Sheng, thank you for being a wonderful intern!

Erin Walker and Amy Ogan, thanks for bringing me into the wolf pack. Ido Roll, thanks for being a great friend, teaching me the ropes, and for being a good sport and letting me prove you wrong.

Moira Burke, Ian Li, Scott Davidoff, Turadg Aleahmad, Aruna Balakarishnan, Johnny Lee, Matt Easterday, Jamie Jirout, Tom Lauwers, Benjamin Shih, Matt Lee, Gahgene Gweon, Gary Hsieh, Martina Rau, Eliane Stampfer, Iris Howley, Nora Presson, Mike Heilman, Laura Halderman, Helen Zhao and the wonderful PIER, PSLC, and HCII students, thank you for sharing not only your insights that greatly improved this work, but also for sharing your friendship, offices, vacations, and stories that greatly improved my life during the past six years.

Finally, the biggest of thanks of all goes to my family for their support and love and especially my mom, to whom this dissertation is dedicated.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

It's an open question whether general strategies for learning exist. Are there instructional techniques that work across domains? Are there strategies that work for students with a multitude of individual differences? If not, what are the boundary conditions? Much work has been done to consolidate highly effective modes of instruction and provide recommendations to instructors. For example, the 2007 IES Practice Guide (Pashler, et al., 2007) highlights seven instructional recommendations that each have empirical support. In this dissertation, I investigate one of these strategies, prompted self-explanation, which is one of only two in the guide determined to have the strongest level of empirical support. I ask whether it is truly domain general as implied by the Practice Guide or whether there are limits to its generalizability. While the vast number of previous self-explanation studies have looked at traditional STEM (science, technology, engineering, and math) domains, I move to the domain of second language grammar learning, a domain that differs greatly from STEM.

Prompted self-explanation is an instructional strategy in which students are asked to provide rules or justifications while problem solving or studying worked examples and has been proven to be an effective instructional strategy for a variety of domains. In early work, Chi (1989) and colleagues examined students learning from worked-example physics problems. In their study, verbal protocols collected during the study phase revealed a correlation between the amount of explanations or justifications that students generated while studying examples and successful problem solving on posttest measures. They observed that students who learned the most used the examples to develop understanding of key ideas while students who didn't demonstrate as much learning on posttest measures used the examples as a reference to look for solutions. Later studies investigated whether prompting students to self-explain would improve learning (Chi, et al., 1994; see recommendation #7 in Pashler et al., 2007). In a study in a biology domain (Chi, 1994), middle school students in the self-explanation condition were prompted to self-explain while reading, while students in the control (unprompted) condition received no explicit prompts to self-explain and, in order to control for time on task, read the material twice. While students in both groups showed significant pre to posttest gains, the gains for the prompted group were significantly larger than those in the control group. Furthermore, a within condition analysis revealed a finding similar to the earlier study. Mainly, students who produced more self-explanations during the learning phase showed greater learning gains compared to students who generated fewer explanations.

Self-explanation prompts have also proven to be effective when added to computer-based tutoring systems. One study within the geometry domain investigated the effects of adding self-explanation prompts to a geometry tutor (Aleven and Koedinger, 2002). In this system, students received feedback on their explanations and could input their answers either by typing the name of a geometry principle or selecting it from a provided glossary. One goal of this study was to evaluate whether the self-explanation effect could be achieved when students have a list of options from which to select their explanations. The tutor was evaluated in a classroom study and results are similar to those of Chi, et al.; students in both the explanation and control conditions showed significant pre to posttest learning gains, with students in the explanation condition showing greater learning gains on transfer measures than those in the practice condition. In this study, transfer items were ones that required declarative knowledge to solve and were of a different form than the problems solved in the tutor. In another study, Atkinson, et al. (2003) prompted students to self-explain by asking them to select a probability principle from a multiple-choice list within a computer-based tutoring system. Their results replicate the Aleven and Koedinger finding that improved learning through prompted self-explanation can be achieved when students select principles from a list or menu. In their study, students who were prompted to self-explain performed better on both near and far transfer tasks than students who were not prompted to self-explain. Near transfer problems were those that were isomorphic to those students saw during tutoring but with different cover stories and values, and far transfer problems were ones in which students had to make adjustments to the problem-solving procedure they learned during instruction. Self-explanation has also been beneficial for learning fractions (Rau, Aleven, and Rummel, 2009), computer programming (Pirolli and Recker, 1994; Bielaczyc, Pirolli, and Brown, 1995) and number conservation (Siegler, 2002). In fact, Roy and Chi (2005) argue:

> Self-explanation is a *domain general* constructive activity that engages students in active learning and insures that learners attend to the material in a meaningful way while effectively monitoring their evolving understanding. (p. 272) [emphasis added]

Despite its vast successes, it is not clear that self-explanation is a general principle that yields better learning in all domains. In fact, with the exception of work done in reading comprehension (McNamara, 2004), little work has been done outside of STEM (science, technology, engineering, and mathematics) domains. In this dissertation, I address the question of whether self-explanation is truly domain general or whether limitations exist. Through a series of randomized and controlled *in vivo* studies (Koedinger, et al., 2009), I investigate whether the self-explanation effect holds when applied to second language grammar learning. Specifically, I built a series of computer-based tutoring systems to teach adult English Language Learners the English article system (when to use *a*, *an*, *the*, or no article at all). I chose this domain for a number of theoretical and pedagogical reasons. With respect to theory building, the article system provides a nice example of a rule or heuristic-driven domain that is quite different from math and science. Unlike basic vocabulary learning of concrete objects where students memorize one-to-one mappings between words in their first and second language (e.g., *chair* in English to *isu* in Japanese), the article system is governed by an underlying rule structure that dictates when each article should be used. While explicit representation of this structure is not needed for successful production (in fact, most native speakers cannot articulate the rules governing the correct article choices they make), there is evidence that providing students with metalinguistic information can be useful for second language learning (Ellis, et al., 2006). Metalinguistic information or feedback often takes the form of providing students the underlying grammar rule or structure for a given construct. For example, instead of simply highlighting a student's error, a teacher could give the student the corresponding rule and identify the features of the sentence or paragraph necessary for applying the rule. Self-explanation is a form of metalinguistic instruction, because by explaining, students identify the features and rules important for making article decisions, and this is the first series of studies to systematically test the effects of prompted self-explanation within the context of second language grammar learning. The article domain was also chosen because of a demonstrated classroom need. Articles are one of the most difficult concepts for second language learners to acquire and article errors can be detrimental to the way a piece of writing is perceived (Master, 1997). I return to the pedagogical importance as well as the specific rules governing the domain in Chapters 2 and 3.

1.1 Self-Explanation as Explicit Instruction

While this is the first series of studies to systematically investigate the effects of self-explanation on second language grammar learning, there is a wide body of literature within second language acquisition that identifies the benefits of explicit instruction, particularly explicit corrective feedback. Ellis, et al. (2006) summarize the findings of 11 studies that compare implicit and explicit corrective feedback. Results consistently show that explicit feedback leads to greater learning advantages over implicit corrective feedback (described below). In addition, two of the studies (Kim & Mathes, 2001; Nagata, 1993) report that when surveyed, students exhibit a preference for explicit over implicit feedback. Explicit instruction was operationalized in a variety of manners across the studies, including direct metalinguistic feedback (e.g., Carroll & Swain, 1993), error identification (e.g., Nagata, 1993), and form-focused instruction (e.g., Lyster, 2004). Metalinguistic instruction is characterized by providing the specific rule or grammar construct for a given situation; error identification includes both explicitly highlighting when an error has occurred as well as highlighting the error and providing the correct response; and form-focused instruction is a broad term that typically involves teaching specific language content. Form-focused instruction is in contrast to a more meaning-based approach where the focus is on successful communication and not on the use of specific grammar forms. Broadly speaking, explicit instruction is characterized by a direct, conscious connection to an underlying structure (N. Ellis, 1994) and may take the form of metalinguistic explanations or explicit correction in which the student's response is identified *and* corrected (R. Ellis, et al., 2006). In contrast, implicit instruction is the process by which the structure is understood "naturally, simply, and without conscious operation" (N. Ellis, 1994, p. 2). Implicit corrective feedback is often delivered in the form of a recast. A recast is a form of feedback in which the learner's incorrect speech is presented back to the student with the errors corrected but not highlighted. For example, if a student says, "Yesterday I *goed to the store", the recast might be, "When you *went* to the store, what did you get?" The focus of the exchange is on the meaning and not the underlying structure (Long, et al., 1998). Explicit feedback is hypothesized to be more beneficial than implicit feedback because it helps students notice gaps in their knowledge and thus encourages repairs.

Since explicit feedback has been shown to be more beneficial than implicit feedback in grammar learning, it reasons that engaging students in explicit reasoning about grammar through the use of prompted self-explanations will also improve learning. It is important to note that in the studies I conducted (details in Chapter 4) *all* of the instructional conditions fall under the broad category of explicit instruction since at minimum they all contain on-demand hints with metalinguistic information and at the most extreme, in the prompted self-explanation condition, the main learning activity is asking students to focus on the metalinguistic details. Thus, in the studies, I will be investigating the relative effectiveness of various forms of explicit instruction.

1.2 Challenges in Generalizing Self-Explanation

As mentioned, the main goal of this work is to investigate whether the self-explanation effect extends to a different type of domain. Namely, is prompted self-explanation beneficial for second language acquisition and, more specifically, for learning rule-based grammar constructs? While there are several differences between STEM domains in which self-explanation has historically been studied and second language acquisition (SLA), one primary difference is the complexity, or number of steps involved in the problem-solving process. (See Chapter 2 for more discussion on the differences between STEM and second language grammar acquisition.) To illustrate, I compare a student identifying the correct article for a sentence to a student identifying the measure for a particular angle in a geometry problem. To successfully choose the article, the student must identify the most important feature or features of the sentence (either implicitly or explicitly) and then map these features to the proper article. In contrast, to determine the measure of an angle in geometry, the student must do more than map features to responses; she must plan and perform calculations. It is important to note that it is not the act of performing calculations that differentiates the two domains but the complexity or number of steps involved in arriving at an answer. In other words, the process for determining article usage is simpler than that for determining measures of angles. However, this simplicity does not mean that learning articles is easier than geometry, simply that the process for determining which article to use for a given instance is less complex than determining the measure of a specific angle. Previous work highlights that the effectiveness of instructional strategies can differ according to the target knowledge (Koedinger, et al., submitted). Specifically, within the context of motor skill learning, Wulf and Shea (2002) argue simpler skills benefit from more challenging learning environments (e.g. spaced not massed practice, infrequent not frequent feedback). They argue that the benefits of spaced practice for simple tasks come from students being able to simultaneously hold multiple items in working memory and compare and contrast the different items. However, this advantage dissipates when students complete complex tasks that individually exceed working memory capabilities, thus preventing beneficial comparisons and contrasts from occurring. Their work reinforces the idea that finding a truly domain general strategy is challenging, and it is perhaps better to align instructional goals with instruction strategy in order to achieve success. Interestingly, others have argued that more complex knowledge components require more complex instructional strategies (Koedinger, et al., submitted).

In addition to the domain differences, there are other hypotheses for why self-explanation may and may not be beneficial for second language learning. According to Chi (2000), self-explanation aids learners through a process by which students generate inferences and then map these inferences to their existing mental models. When discrepancies arise, students make appropriate adjustments to correct their models. The Cascade model (VanLehn, Jones, and Chi, 1992) argues that when students self-explain, they identify and fill gaps in their knowledge. This process enables better declarative knowledge of the domain and provides multiple strategies by which students can solve subsequent problems. Self-explanation may help students learn the English article system by highlighting rules or concepts that are not in the students' models and helping them to develop a more complete representation. Self-explanation may also be beneficial by encouraging feature focusing. Asking students to provide explanations facilitates deep processing of the material and encourages students to identify and attend to relevant features of the sentence and disregard those that are not important. However, the process of self-explaining is likely to take time. In second language learning, teachers and students are interested in error-free production and care less about whether students are able to articulate declarative representations of the rules. Perhaps spending time to acquire these explicit declarative representations may prove to be too time consuming, and students would be better served with a simpler, practice-only environment. While self-explanation has been largely successful in other domains, there is evidence that it may be inefficient at times. In their study investigating the effects of molar and modular worked examples for learning probability word problems, Gerjets, et al. (2006) found no learning advantage for prompting students to self-explain over a worked-example condition that provided elaborated explanations. Furthermore, timing results showed that prompting students to self-explain took significantly more time. The authors hypothesize that the results could be due to several design decisions: a split-attention effect as a result of the explanations being presented on a separate page, students not receiving feedback on explanations because they were typed versus selected from a provided list, and students being overwhelmed by the number of self-explanation prompts and perceived redundancy of the explanations. Similarly, Matthews and Rittle Johnson (2009) found no advantage for self-explanation when conceptual instruction was provided and argue that under certain circumstances, self-explanation may be an inefficient strategy.

The studies in this dissertation highlight the importance of evaluating instructional techniques both in terms of effectiveness and efficiency. Students do not want to study longer than necessary, and teachers have an ever-growing curriculum that they need to cover. Thus, it is critical when designing classroom interventions, to take into account the temporal aspects of learning and create instruction that achieves the maximum learning benefit relative to the amount of time invested. Several studies have succeeded at using computer-based instruction to improve instructional efficiency. In a study by Lovett, et al. (2008), students completed an 8-week online version in lieu of a 15-week traditional classroom-based statistics course. Students in the online version were as good or better at immediate and delayed post-tests but only spent half as much time on the instruction. In the studies described in Chapter 4, I explore both the effectiveness and efficiency of prompting students to self-explain while learning second language grammar.

In this dissertation, I discuss the results of a series of tightly controlled, classroom-based studies (*in vivo* experiments) that were made possible through collaboration between the Pittsburgh Science of Learning Center (www.learnlab.org) and the University of Pittsburgh's English Language Institute, an intensive language program for adult English language learners. I worked with classroom teachers to design computer-based tutoring systems that fit within the existing curriculum, enabling students to participate as part of their regular grammar classes. In the following chapters, I describe the design and evaluation of several computer-based tutoring systems that address the question of whether the benefits of self-explanation extend to second language grammar learning. In order to build effective systems, I first used an iterative approach to build a knowledge component model for second language learners learning the English article system (when to use *a*, *an*, *the*, or no article at all) (Chapter 3). I then used this model in the design of various computer-based tutors that were evaluated in classroom studies (Chapter 4). In a pilot study, I explored different modalities of self-explanation (free-form responses versus menu selections). Study 1 examined the effects of tutored practice versus self-explanation alone on procedural and declarative knowledge acquisition and learning efficiency. Studies 2 & 3 introduced additional instructional strategies (Study 2 – analogical comparisons, Study 3 – worked examples) as further points of comparison, and finally Study 4 compared an adaptive self-explanation tutor to tutored practice on normal learning and long-term retention measures. In Chapter 5, I discuss the results of cross-study analysis of learning curves in order to examine the learning process, and in Chapter 6, I examine effects of individual differences (prior knowledge and students' native languages). Finally, in Chapter 7, I discuss the conclusions and contributions from this work. Broadly speaking, this work demonstrates the importance of tempering generalizations about instructional strategies with more precise claims and supporting theory for the specific type(s) of knowledge acquired. More generally, these studies demonstrate how educational technology and human-computer interaction techniques can be used to run tightly controlled and finely instrumented studies within existing courses.

Chapter 2: The English Article System

In this work, I investigate whether the benefits of self-explanation extend beyond STEM domains, specifically whether it is an effective instructional technique for adult non-native English speakers learning the English article system. The instruction teaches students when to use *a*, *an*, *the*, or no article at all. I chose to focus on this domain because of the importance of acquiring article use skills, the rule-based structure of the article system, and second language acquisition theory that suggests articles may be particularly well suited for a rule-based instructional approach as described below. I first describe the domain and the observed difficulty that students have in acquiring it. I then highlight the differences between second language acquisition and STEM domains as well as the differences between first and second language acquisition.

Even though article errors often do not hinder communication, their presence, especially in written documents, signifies a non-native speaker and can have detrimental effects with respect to how the piece is perceived (Master, 1997). Grammar errors can also cause writing to become "unintelligible", "irritating", or both (R. Ellis, 1994). The article system is one of the last grammar constructs for English as a Second Language (ESL) students to acquire (Master, 1997), and errors are commonly seen in student writing. In an error analysis conducted with 75 English Language Learner writing samples, article errors were the most common error type and represented 16.7% of the total errors. Essays were collected as part of normal classroom activities (i.e., a combination of both in-class and homework assignments) and were subsequently coded by an experienced ESL instructor. Each error was tagged and coded with the specific grammar rule that was violated. This analysis confirmed the claims made by teachers and students: article acquisition is very difficult for students and thus a successful tutoring system would be a helpful addition to a student's course of study. As DeKeyser (2005) emphasizes, articles are difficult to learn because the English system is frequently different from that used in a student's first language, and the meaning articles express is often subtle and difficult to infer. However, one of the advantages of working within such a difficult domain is that even students who are proficient in other areas of the language have yet to master the article system. This context allows for a greater exploration of the instructional design space and enables the variables important for article instruction to be isolated. For example, target students already have relatively large vocabularies and a solid understanding of basic grammatical structures, giving more flexibility in the design of the instructional conditions.

2.1 Student Population

When designing instruction, in addition to having a solid understanding of the target domain, it is important to understand the target student population. My target students are adult English language learners who are learning English for academic or professional purposes with at least an intermediate level of proficiency (at least a TOEFL score of 400). Previous research suggests that students below an intermediate-level of proficiency do not have enough information (e.g., vocabulary, grammar structures, etc.) to acquire skill in the English article system (Master, 1997). The majority of the students have had formal education of English as a foreign language for many years (middle school and high school). In addition, the students enrolled in intensive language programs are often highly motivated. They have made large personal and financial commitments by moving to the United States to further their study of English. Often their goals are to enroll in an undergraduate or graduate degree program in an English-speaking country or to get a job that requires near-fluent English skills. These students are different from people learning English for conversational or travel purposes. Outside of a professional context, correct article usage rarely interferes with communication and thus would it would be inappropriate for students desiring that level of mastery to spend much time on articles when time would be better spent in another area, such as vocabulary building.

2.2 STEM vs. Second Language Acquisition

While self-explanation shows great promise in increasing learning, as mentioned, most studies have dealt with math and science domains and relatively little work has been done in areas like second language acquisition. There are inherent differences between math and science and second language acquisition that may affect the success of self-explanation in this new domain. One key difference lies in the pedagogical goals of the domains. In math and science, students are often expected to be able to solve problems and explain the underlying principles. For example, in geometry, students must know not only how to determine the angles of a triangle but also how to explain why their answers are correct. However, the marker of a successful language learner is fluency, and knowing when and how to use a particular construct is much more important than knowing why. In fact, most native English speakers have little explicit knowledge of grammar rules yet rarely make mistakes. Another difference lies in the presence and absence of exceptions to rules. In math and science, there are no exceptions. Provided that the proper conditions are met, a given rule will always apply. However, there are frequent exceptions to grammar rules. For example, one rule listed in an English as a Second Language grammar book states that no article should be used before the name of a disease (Cole, 2000); for example, He has (no article) diabetes. However, many exceptions exist (e.g. He has *the* flu, She has *a* cold.) Perhaps encouraging students to focus on rules and features of the sentence, which may sometimes prove to be unreliable, is not an effective learning strategy. Furthermore, it is an open question whether it is worth the instructional cost of acquiring declarative knowledge for this domain. One advantage of declarative knowledge is that it provides an additional path by which a problem can be solved. For example, in the sentence "Yesterday, I bought a car. Today, _*the*_ car broke," students could know to use "the" because it implicitly "sounds correct" or by correctly applying the rule "if a noun has already been mentioned, use *the*." However, it is also possible that even if a declarative knowledge path can be established, it may not be preferred due to the extra processing time required. Since language fluency is measured using time and accuracy, a successful strategy must optimize both.

2.3 First Language vs. Second Language Acquisition

As any adult learning a second language can attest, there are many differences between acquiring one's first language as a child and learning a second language as an adult. A number of theories exist as to why these learning processes are drastically different. For example, proponents of the critical period argue that after a certain age, the brain changes in a way that makes learning languages challenging (Lenneberg, 1967). Others claim that the difference in learning rate and success is due to socio-cultural differences. For example, adult second language (L2) learners are often less immersed in the L2 environment and their conversations with native speakers contain fewer scaffolds or salience markers than infants acquiring their native language (Schumann, 1979). A more recent theory situates itself in the cognitive principles of transfer and suggests that students learning a second language have difficulty because of the associative learning of form-meaning relationships (Ellis & Sagarra, 2010). While the reasons for differences between L1 and L2 acquisition may be debated, for the purposes of this work, it is simply important to note that they exist. Specifically, the three key differences for this work are: learning outcomes, the age of the learner, and the role of explicit instruction. Perhaps the most striking difference between first and second language acquisition is the probability of success. In the absence of impairments, everyone is able to successfully acquire a first language. In contrast, fluency in a second language is far less certain. Another key difference is the age of learner. Despite the name, "second language" refers not to the order in which a language was acquired but describes any language that is learned after the critical period (usually defined as after puberty) (Lenneberg, 1967). This milestone is important due to the many developmental changes, both cognitive and metacognitive, that occur during and after early adolescence. Finally, the role of explicit instruction varies greatly between first and second language learning. Children learning their first language do not need explicit instruction, and often ignore it when it is provided (Mitchell & Myles, 2004). However, explicit instruction in second language classrooms is not only common but some SLA theories suggest that it is required for certain grammar constructs to be noticed and learned (Williams & Evans, 1998).

Chapter 3: Knowledge Component Model Building and Refinement

3.1 Knowledge Component Model Building

While it may seem obvious that in order to learn a skill, one must practice it, deciding *what* students need to repeat or practice during a course of instruction is not always an obvious choice. This challenge is especially true for domains like the English article system where the rules or knowledge components may be unclear or unknown. In the Knowledge-Learning-Instruction (KLI) framework (Koedinger, Corbett, and Perfetti, submitted), a knowledge component (KC) is defined as "an acquired unit of cognitive function or structure that can be inferred from performance on a set of related tasks." The term KC is encompassing of any small, but lasting mental change that yields consistent performance differences over time and task contexts, and it includes rules, principles, concepts, facts, schemas, mental models, etc. Thus, a knowledge component model is a collection of the KCs that describe the requisite knowledge for a given domain and map to tasks that require that knowledge. In this chapter, I describe the iterative process of defining a knowledge component model for second language learners learning the English article system. I first present the initial "behavioral" KC model (the answer model) and the empirical results of using this model for instruction and assessment alignment. While the learning results of this early study were very poor, the data collected as part of the study highlight many limitations of this simple model, and suggest that a more detailed, feature-based "cognitive" model is more appropriate. I then discuss the follow-up study in which instruction and assessments were aligned based on the second iteration of the KC model. In this study, I validated the new model by conducting a within-subjects study in which I controlled for whether students received tutoring or not for various KCs. Furthermore, the data from the study were used for a third and final iteration of the knowledge component model, and this final model was used in the design of all subsequent studies.

3.2 Knowledge Component Model: Iteration and Validation #1

The first iteration of the knowledge component model (*The Answer Model*) was based on the desired behavior or answer (*the*, *a/an*, and no article) that students should provide for a given sentence and thus contained only three knowledge components (Table 1).

Table 1. Article Book Knowledge Component Model

KC	Answer	Description
KC1	the	All definite nouns
KC2	a/an	All indefinite, singular nouns
KC3	no article	All plural or non-count indefinite nouns

While admittedly simplistic, this knowledge component model was based on basic categorization and organization used in textbooks (e.g., Fuchs, et al., 2003) and also is at approximately the same grain size as Bickerton's semantic wheel (1981). Bickerton's Semantic Wheel (Figure 1) takes into account the perceived knowledge of both interlocutors. Each noun is labeled as having a specific referent or not (SR+/-) and for whether or not the hearer has assumed knowledge of the item (HK+/-). While this model can be helpful in identifying why students are making specific errors (e.g., Chinese users may confuse HK+/- (Lu, 2001)) it is not very helpful pedagogically since one quadrant may map to several answers. Thus, even if students are able to identify the correct noun type and quadrant, they may still not know which article to use.

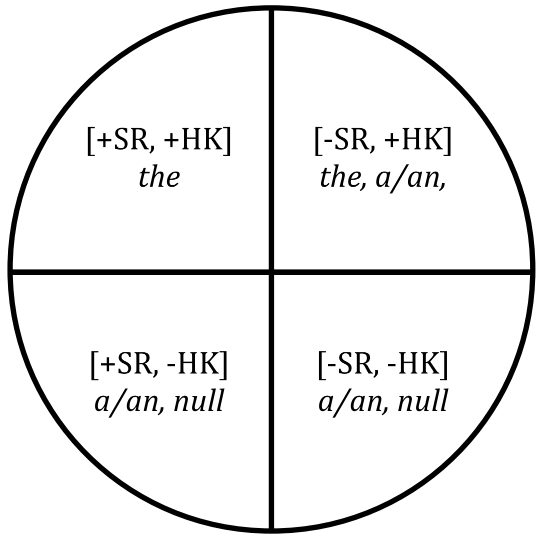


Figure 1. Bickerton's Semantic Wheel

In order to assess the validity of this early answer-based KC model as well as the effectiveness of an early article tutor design, I conducted a lab study. In this study, students edited entire paragraphs by completing cloze (fill-in-the-blank) tasks and selecting the article (a, an, the or no article) that best completed the sentence (See Figure 2). Participants received immediate feedback on their selection (answers turned green if they were right, red if they were wrong), and participants had access to on-demand hints to prevent floundering. The paragraphs were selected because they were at an appropriate reading level and because they contained several instances of each of the target knowledge components (i.e., several uses of "a", "an", "the" and no article). Participants began the session by completing a short pretest, which employed the same interface as the tutoring phase, but participants did not receive feedback on their selections. Participants were then tutored on three paragraphs (42 total article selections), and instruction was designed such that students would have several practice opportunities with each of the target knowledge components (Table 2). Finally, students completed a posttest, which was counterbalanced with the pretest. In total, there were four assessment paragraphs (two easy and two difficult forms adapted from intermediate and advanced textbooks respectively). Each student completed one easy and one difficult assessment, and forms were counterbalanced across pretest and posttest.

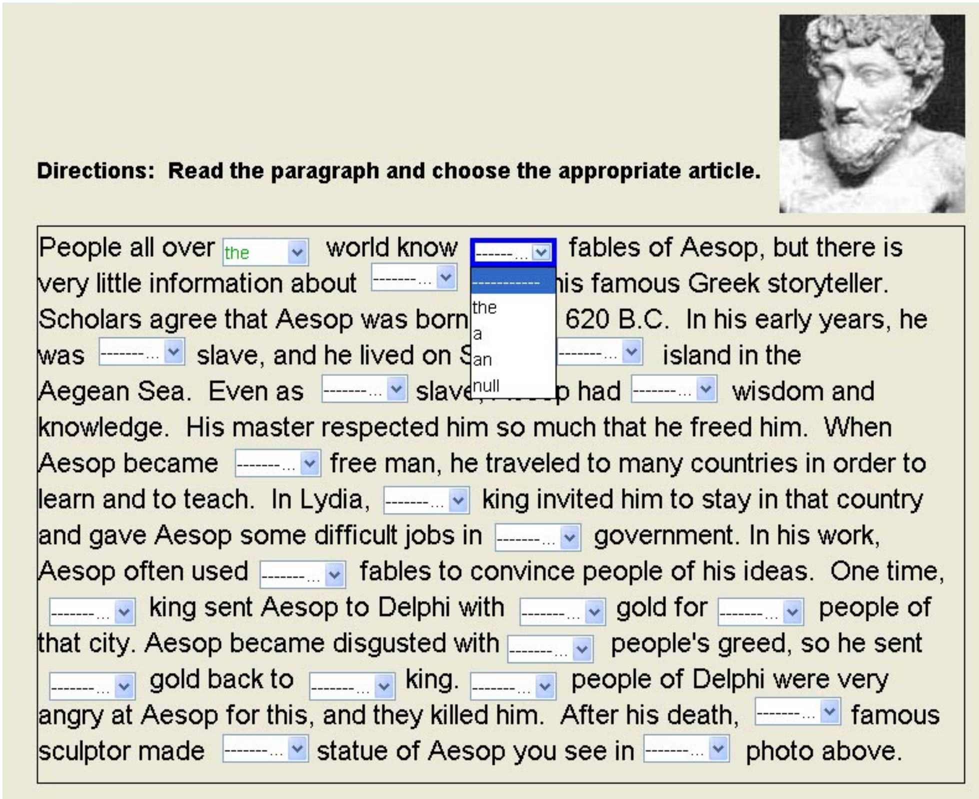


Figure 2. In an early version of the article tutor, students selected the article that completed the sentence within the context of a larger paragraph.

All test and tutor interfaces were developed using the Cognitive Tutor Authoring Tools (Aleven, et al., 2009) and deployed online via Java WebStart. All student actions (answer attempts, hint requests) were logged and time stamped.

Participants

Adult, non-native English speakers (N=21) were recruited to participate in the lab study. Prior to taking the pretest and using the tutor, participants were given instruction on how to use the interface but no explicit grammar instruction. If students had questions while using the system, they were instructed to request a hint through the interface. For questions that arose during the pre and posttests, participants were simply told to do their best and all questions were answered during the debriefing session at the completion of the study.

Table 2. Alignment in early tutor designs. Instruction was designed to give students several practice opportunities with each of three knowledge components.

Knowledge Component	Tutored Paragraph #1	Tutored Paragraph #2	Tutored Paragraph #3	Total
KC 1: the	4	5	4	13
KC 2: a/an	2	3	4	9
KC 3: no article	10	5	5	20

Learning Results

Initial learning results were disappointing and showed no learning gains. Students did no better on the posttest than they did on the pretest (71.1% pretest vs. 69.8% posttest, $t(1,20)=0.29$, $p = 0.77$). These results could be driven by a number of factors: (1) students were not presented with enough learning opportunities (insufficient practice), (2) the instruction was effective but the assessments failed to capture student learning (poor KC alignment), or (3) a combination of both these hypotheses.

While the instruction content was adopted from popular and effective grammar textbooks and designed to give students multiple opportunities to practice each knowledge component (9 uses of "a/an", 13 uses of "the", and 20 uses of the zero article), I realized that one critical reason for why aligning based on answer was perhaps too simplistic was that each article use may have several grammatical features behind it. For example, the article "the" can be used for many reasons, including "the noun has already been mentioned" (e.g., Yesterday, I bought a car; today **the** car broke); "the noun is unique" (e.g., **The** moon looks very large tonight); "the noun is modified by an ordinal number" (e.g., Noah was **the** first person to come to the party), etc. To test whether instruction and assessment should be aligned at this deeper, feature level rather than the answer level. I developed a second knowledge component model (*The Feature Model*) based on an ESL textbook (The Article Book, Cole 2000). This model contains 66 rules for determining article usage (See appendix for complete list). When the instruction and assessment items were recoded based on the features of the sentence, out of the 66 potential rules, the sentences used in the tutor and assessments were classified using only nine (Table 3).

Table 3. Original Tutor and Assessment Items recoded using the *Feature Model*

KC	Answer	Feature
KC1	a/an	General Singular: Use "a" or "an" when a singular count noun is indefinite. (e.g., I want a car.)
KC5a	no article	General Non-count: Do not use an article with a non-count noun that is indefinite. (e.g., At a birthday party, people usually eat (no article) cake.)
KC5b	no article	General Plural: Do not use an article with a plural noun that is indefinite. (e.g., (no article) Scientists argue that global warming…)
KC12	the	Already Mentioned: Use "the" when the noun has already been mentioned. (e.g., Yesterday, I bought a car; today, the car broke.)
KC13	the	Already Known: Use "the" when the noun is already known. (e.g., I know that restaurant. The chef is a friend of mine.)
KC14	the	Prepositional Phrase: Use "the" when the noun is made definite by a prepositional phrase. (e.g., The tree in front of my house is oak.)
KC23	the	Ordinal Number: Use "the" with ordinal numbers and ranking words like "next" and "last". (e.g., Noah was the first person to arrive at the party.)
KC25	the	Classes: Use "the" when generalizing about an entire class of animals. (e.g., Many animals depend on the mosquito.)
KC33	the	Unique for All: Use "the" when the noun is the only one that exists. (e.g., The moon looks very large tonight.)

Furthermore, recoding the original stimuli with the feature model highlights areas where the instruction and assessment were misaligned (Table 4). Namely, while KC1, KC5a, and KC5b were all seen frequently (with a total of 8 or more learning opportunities) during training, the other KCs were seen rarely. This new model suggests that not all the knowledge components were seen enough times during tutoring to affect a change in performance.

Table 4. Recoding the stimuli with the feature model highlights discrepancies in the alignment between assessment and instruction.

Test Form	Form 1 (easy)	Form 2 (difficult)	Form 3 (easy)	Form 4 (difficult)	**Total Assessment Items**	Tutor Problem 1	Tutor Problem 2	Tutor Problem 3	**Total Tutor Items**
KC1 (singular, generic)	5	2	1	1	**9**	2	3	4	**9**
KC5a (noncount, generic)		6	2	6	**14**	3	2	3	**8**
KC5b (plural, generic)		3	1	8	**12**	7	3	2	**12**
KC12 (already mentioned)	7		4		**11**		1		**1**
KC13 (already known)	3		2	4	**9**		2	1	**3**
KC14 (prep phrases)		4	1	1	**6**	1		1	**2**
KC23 (ordinal numbers)	1		1	1	**3**	1	1		**2**
KC25 (class)		1			**1**				**0**
KC 33 (only one)		1	1	3	**5**	2	1	2	**5**

Table 5. IRT results: For each assessment

	Learning Expected = No	Learning Expected = Yes
High Frequency	16 (45.7%)	19 (54.3%)
Low Frequency	24 (68.6%)	11 (31.4%)

More specifically, if this model is correct, one might expect to see learning of KCs 1, 5a, and 5b but not KCs 12, 13, 14, 23, 25, and 33. In order to estimate the effect of tutoring, I used an IRT model to calculate item difficulty parameters at pretest and posttest (For results, see Appendix 3). Under the hypothesis, items that assessed knowledge components that were *frequently* seen in the tutor should have a lower difficulty parameter when seen at the posttest than at the pretest. On the contrary, items for which the knowledge component was not seen frequently during tutoring should show no difference at posttest or perhaps be harder at the posttest in the event that students are developing a bias towards tutored knowledge components and against non-tutored items (for example, if students frequently get the item wrong when selecting "the" during tutoring, they may be less likely to select "the" on the posttest). Thus, I first calculated difficulty parameters for each item using only the pretest data and then repeated the analysis using only the posttest data. Using these estimates, I then calculated for each item whether the model predicts that learning would occur. Namely, if the posttest difficulty estimate is less than the pretest estimate, learning is expected; otherwise, no learning is predicted (Table 5). In other words, since the test forms were counterbalanced, if an item receives a lower difficulty estimate on the posttest than on the pretest, one can attribute the reduction to student learning. However, if an item has a higher or equal difficult estimate on the posttest than on the pretest, it can be assumed that that item was not learned during the course of instruction. Chi-squared analysis shows that according to the model, high frequency items are more likely to show learning than low frequency items (χ=3.73, p=0.05)

This pattern was further validated by a linear mixed model analysis in which I grouped data by student and used whether the student got the item correct or not as the dependent variable and test time (pretest vs. posttest) and frequency within the tutor (less than seven learning opportunities equals low frequency, seven or more learning opportunities equals high frequency) as factors. This reveals an interaction between test time (pretest vs. posttest) and item frequency (low vs. high). Namely, there was a pre to posttest gain for items that were seen more often in the tutor (high frequency) but not for low frequency items ($F(1, 725.9) = 3.87$, $p = 0.049$) (Figure 3). These results were promising and suggested the *Feature Model* was a useful knowledge component model for this domain. Thus, in the next round, I redesigned the tutoring system in order to deliver instruction that was aligned using the new model and empirically validated the model in a classroom study.

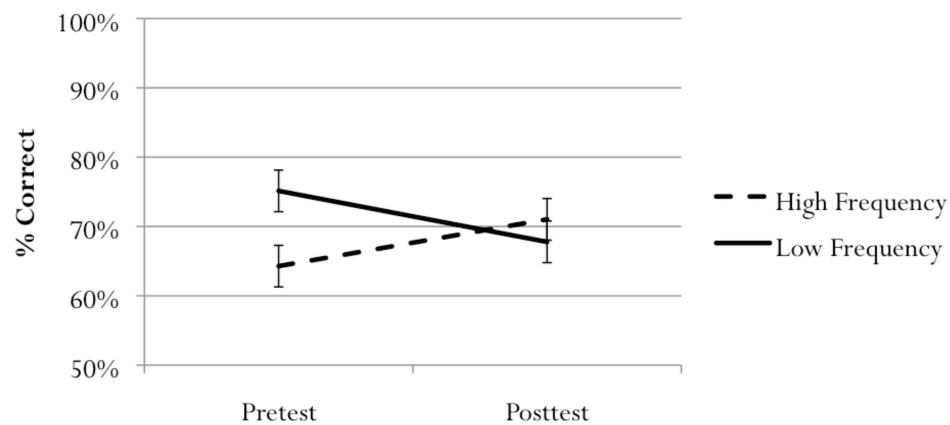


Figure 3. Interaction between item frequency and test time. Knowledge components that were seen frequently in the tutor showed an increase in performance on the posttest compared to the pretest. Knowledge components that were not seen frequently during tutoring showed no learning gains.

3.3 Knowledge Component: Iteration and Validation #2

In order to develop instruction that was aligned using the *Feature Model*, I first changed the format of the instruction. In the first iteration of the tutor, students completed sentences that were part of a larger paragraph; however, in order to have more control over the type and frequency of knowledge components that students saw, I changed from paragraph-level instruction to sentence-level instruction. Thus, in the new versions of the tutor and assessments, students were presented with individual cloze (fill-in-the-blank) sentences and chose the article that fit best (Figure 4). As in the previous lab study, during tutoring, students continued to receive immediate feedback on their selections and had access to hints.

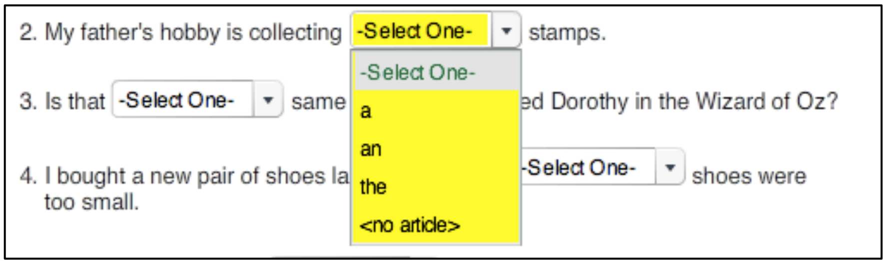


Figure 4. In the new design of the Article Tutor, students select articles to complete individual sentences rather than complete paragraphs.

In order to validate the new knowledge component model, I conducted a within subjects study in which students received tutoring on a subset of the knowledge components on which they were tested. This study was part of the larger Study 1, which is discussed in more detail in Chapter 4. To validate the knowledge component model, I developed three sets of four knowledge components chosen from the 66 total KCs (Table 6). The knowledge components were chosen in collaboration with ESL instructors.

Table 6. The target knowledge components were placed in one of three groups. All students were tutored on KCs in Group A but were randomly assigned to receive tutoring on KCs in Group B or C. All students were assessed on all 12 KCs.

Group A	**Group B**	**Group C**
Use "no article" with plural, generic nouns.	Use "a" or "an" to mean "for each" or "per".	Use "the" when the noun is made definite by a prepositional phrase
Use "a" or "an" for single letters and numbers.	Use "an" when a singular count noun is indefinite and the article is followed by a vowel sound.	Use "the" when the noun has already been mentioned.
Use "the" with the word "same".	Use "the" with ordinal numbers and other ranking words like "next" or "last"	Use "a" when a singular count noun is indefinite and the article is followed by a consonant sound.
Use "no article" with non-count generic nouns.	Use "the" when the noun that follows is already known.	Use "the" when the noun is made definite by an adjective clause or an adjective phrase.

All students were tutored on the knowledge components in Group A; however, students were randomly assigned to either study KCs from Group B or Group C. Since only Groups B and C were randomly assigned, only these KCs are included in the following analysis. I hypothesized that students would demonstrate significant improvement only on the knowledge components on which they received tutoring. In order to test this hypothesis, I conducted a mixed model analysis, grouping by students and using whether or not students got the answer correct or not as the binary dependent variable. Independent factors were test time (pretest vs. posttest) and knowledge component category (tutored vs. non-tutored). Results show a significant interaction between test time and tutor status ($F(1,22.17)=21.19$, $p < 0.001$) (Figure 5); namely, students performed better on the posttest (compared to the pretest) for knowledge components that were tutored but not for those that were not tutored.

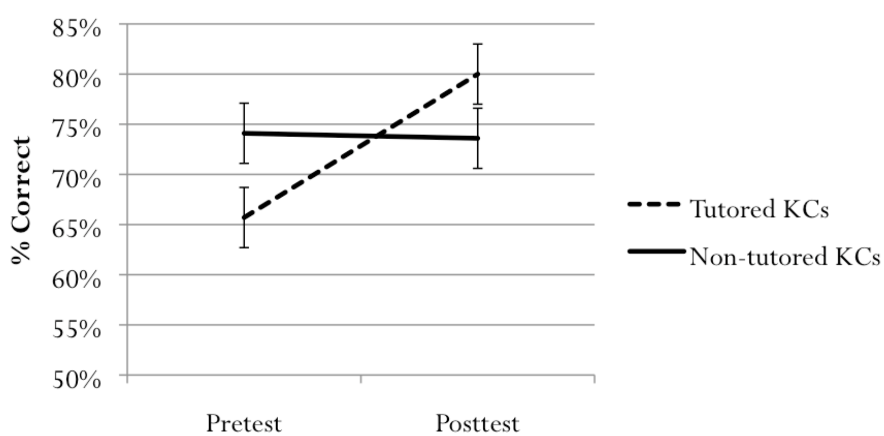


Figure 5. As hypothesized, students demonstrate significant pre to posttest learning gains only for knowledge components on which they receive tutoring.

This analysis suggests that the new *Feature Model* is a helpful model for aligning instruction and assessment. However, the data collected from the study was once again used to refine the knowledge component model. In particular, it was hypothesized that certain knowledge components could be combined to make a more parsimonious model. For example, in the model, there are two knowledge components for singular, generic nouns (KC2 and KC3). These two KCs differed in that one is for nouns that start with a consonant (KC2) sound and the other is for nouns that start with a vowel sound (KC3). I hypothesized that the more important of the features was the singular and generic distinction and not the phonetic distinction. Thus, I created a new knowledge component model combining these two KCs into a single KC (If a noun is singular and general, use "a" or "an".) I then compared the original (12 parameter KC model) to the revised (11 parameter model) using BIC scores as the metric. BIC scores were calculated using the Pittsburgh Science of Learning Center's DataShop modeling tools (Koedinger, et al., 2010) and revealed that the more parsimonious 11-parameter model (BIC=8132.98) was a better model for the data than the original 12-parameter model (BIC=8139.35). Therefore, in subsequent studies I used the simpler in designing instruction and assessments.

In this chapter, I described the iterative process of developing a knowledge component model for the English article domain. The final model was initially seeded with theoretical knowledge from second language acquisition and refined using empirical data. While more iterations of the model would likely lead to improvements (and some are discussed in Chapter 5), I'm reminded of the quote by statistician George Box (1979), "All models are wrong, but some are useful." While the current knowledge component model may not be 100% correct, it has proven to be a useful tool for aligning instruction and assessment.

Chapter 4: Classroom Studies

In this chapter, I describe the series of in vivo classroom studies that I completed in order to empirically test whether prompted self-explanation is a beneficial strategy for English article acquisition. After an initial pilot study in which I compared two modalities of self-explanation, I then investigated the use of practice and self-explanation alone (Study 1). In Studies 2 and 3, I added other, related instructional conditions: analogical comparison (Study 2) and worked examples (Study 3). Finally, in Study 4, I built and evaluated an adaptive self-explanation tutor that prompted students to explain only when evidence of their prior knowledge for a given knowledge component was low.

4.1 Pilot Study: Exploring Different Forms of Self-Explanation

I first conducted a pilot study to test the feasibility of a classroom study, evaluate the instruction and assessment materials and investigate two modalities of prompted self-explanation. I created two versions of a self-explanation tutor (menu-based and free-form response) and a tutored practice (practice-only) condition. In the practice-only tutor, students selected an article (*a*, *an*, *the*, or *no article*) from a drop-down menu to complete the sentence (See Figure 6). The practice-only tutor served as an ecological control even though, due to the addition of immediate feedback and hints, it was arguably better than current classroom practice. To investigate the effects of self-explanation, I 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. In the free response self-explanation tutor, students first selected the correct article and were then asked: "Why is that the answer? Which rules or features did you use to make your choice?" Students typed their responses in textboxes. All answers were accepted, and no feedback on their explanation was given (Figure 7). Students had access to hints to aid with both the article selection and self-explanation steps (Table 7). The hints identified the relevant features of the sentence and then provided the rule that dictates which article should be used. One of the potential disadvantages of the free response method of self-explanation was that the system could not easily provide feedback to students on their explanations. However, if students were to select a rule or explanation from a given list, as has been used in studies of self-explanation in geometry (Aleven & Koedinger, 2002) and probability (Atkinson, Renkl, and Merrill, 2003), the tutor could provide relevant feedback and ensure students' explanations are correct before continuing. In the probability tutor, students selected from a list of four rules/principles. In a similar fashion, students using the menu-based article tutor first selected the correct article and then chose an explanation for their article choice from a drop-down menu (Figure 8). Students received immediate feedback and again, identical to the free response self-explanation tutor, had access to hints. Due to limited classroom time, I chose to focus on a subset of the knowledge components for the domain. With the help of the students' classroom teachers, we selected 8 KCs to cover in the tutor. The 8 KCs included both very frequent rules that we expected to be familiar to students to as well as rare rules that we expected to be new to students. See appendix for list of knowledge components covered in each of the studies.

Table 7. Example hints used in the pilot study.

Target Sentence: Last week, I bought a TV. Today, ___ TV was stolen.

	Hint text for practice (article selection) tasks	Hint for the self-explanation tasks.
1	TV has already been mentioned.	"TV" was mentioned in the first sentence.
2	When a noun has already been mentioned, use "the".	Since "TV" was already mentioned, it is definite.
3	Please select "the" from the highlighted menu.	Please select "The noun has already been mentioned" from the highlighted menu.

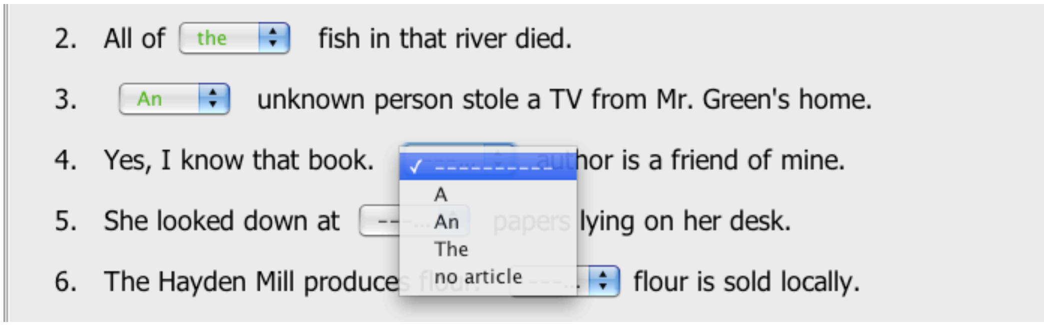


Figure 6. Practice Tutor. Using the practice tutor, students selected the article that best completes the sentence. Students received immediate feedback and had access to hints.

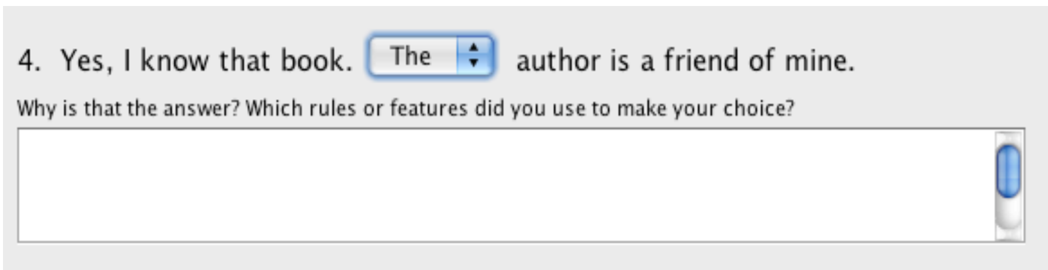


Figure 7. Free-form Self-Explanation. Using the free-form self-explanation tutor, students selected the article that best completed the sentence and then generated and typed an explanation. Students received immediate feedback on their article choice but no feedback on their explanation, but had access to hints for both steps.

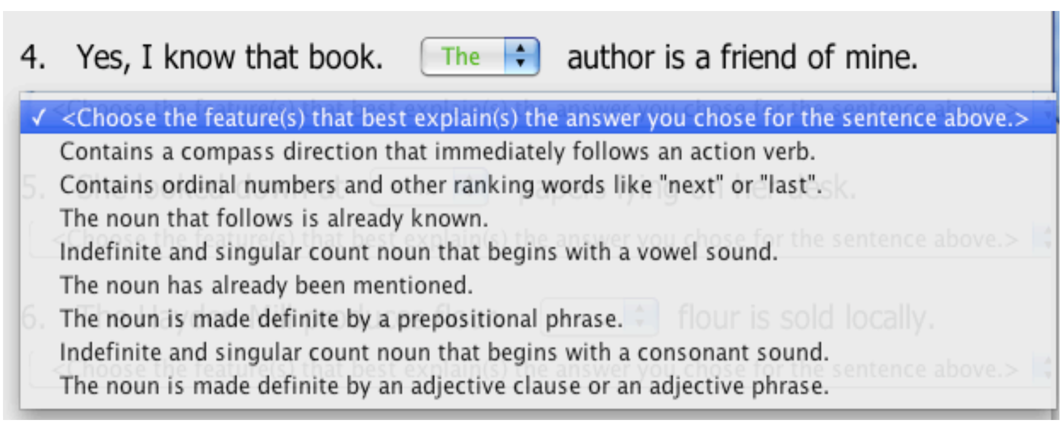


Figure 8. Menu-based Self-explanation. In the menu-based self-explanation tutor, students selected the article to complete the sentence and then selected the explanation. They received immediate feedback and had access to hints for both steps.

4.1.1 Pilot: Procedure

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 spoke a variety of first languages. For this pilot study, students in the intermediate (n=15) and high-intermediate (n=43) courses completed the tutor, pretest, and posttest as a one-day in-class activity, while students in the advanced course (n=5) completed them as a homework assignment. All students completed a computer-based pretest 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 then randomly assigned to a tutor condition. In an attempt to keep time-on-task equal, students in the no self-explanation condition completed three times as many article selection tasks (84 sentences) as in the self-explanation tutors (28 sentences). The decision to have students complete more sentences was made after pre-pilot data showed that completing 28 article-selection (practice) items took about one third the time as completing 28 matched self-explanation items. Finally, students completed an immediate posttest (isomorphic to and counter-balanced with the pretest) and a survey.

4.1.2 Pilot: Results

The main goals of the pilot study were to (1) test the classroom feasibility of the study and determine whether the instruction time was sufficient to observe learning, (2) insure that the assessments were aligned with the instruction, and (3) investigate the differences between free-form (answers typed without feedback) and menu-based (answers selected with feedback) self-explanation. The assessment items were divided into two categories: procedural fluency items (the article selection tasks) and verbal declarative knowledge items (the explanation selection tasks). A repeated measures ANOVA with score on the procedural fluency pre and posttest as the dependent variable revealed a significant main effect for test time ($F(1,60) = 40.0$, $p < 0.001$) indicating a significant pre to posttest gain, and a marginal interaction between test time and condition (F(2, 60) = 2.78, p = 0.07) (Table 8). This possible interaction may result from a lower learning gain on the part of the menu-based self-explanation group. However, given the bigger group differences were at the pretest rather than posttest, the apparent interaction may simply be statistical noise (an "unhappy randomization") in this small-sample pilot study.

Table 8. Pretest, Posttest, Gain and Tutor Time by Condition. Overall, there was an effect of test-time (pretest to posttest) and a marginal interaction between test time and condition. In addition, students in the menu self-explanation (SE) with practice conditions (Figure 8) completed the instruction significantly faster than those in the free-form self-explanation (SE) with practice condition (Figure 7).

Condition	N	Pretest (SD)	Posttest (SD)	Average Normalized Gain (SD)	Instructional Time* (SD) (minutes)
Tutored Practice	21	60.4% (13.5)	74.4% (11.5)	34.4% (27.4)	15.2 mins (6.8)
Menu SE with Practice	21	72.6% (20.0)	78.0% (14.9)	19.7% (32.0)	12.9 mins (6.0)
Freeform SE with Practice	21	58.9% (14.3)	73.5% (11.8)	35.0% (32.4)	18.5 mins (7.8)
Averages	63	64.0% (17.1)	75.3% (12.8)	29.7% (31.0)	15.5 mins (7.2)

A similar analysis for the declarative knowledge assessments showed neither pre to posttest learning (F(1,60) = 1.32, p = 0.26), nor an interaction between test time and condition (F(2,60) = 0.29, p = 0.75) (Table 9). These results were not consistent with prior results (Aleven & Koedinger, 2002) nor with our subsequent results (see Study 1 below) – these other studies found explanation gains for both tutored practice and self-explanation groups.

Table 9. Explanation choice pretest, posttest and gain scores by condition.

Condition	N	Explain Pretest (SD)	Explain Posttest (SD)	Average Normalized Gain (SD)
Tutored Practice	21	39.3% (19.9)	41.1% (26.9)	-7.4% (48.9)
Menu Explanation with Practice	21	52.4% (25.8)	53.6% (29.1)	6.8% (37.8)
Freeform Explanation with Practice	21	37.5% (27.7)	42.9% (30.3)	8.5% (31.7)
Average	63	43.1% (25.2)	45.8% (28.8)	2.6% (40.1)

To understand learning as a cognitive process, it is important to understand not only how learning changes performance but also the time course in which it occurs. Thus I investigated the efficiency of learning induced by the different instructional conditions. There was a significant difference between the conditions in the amount of time it took students to complete the tutors ($F(2, 60) = 3.51$, $p = 0.036$). Post-hoc Tukey HSD tests revealed that students using the menu self-explanation tutor completed the instruction the fastest ($M = 12.9$ minutes, $SD = 6.0$) but not significantly faster than those in the tutored practice condition ($p = 0.541$, $M = 15.2$ minutes, $SD = 6.8$), and the practice tutor was not significantly faster than the free-response tutor ($p = 0.266$, $M = 18.5$, $SD = 7.8$) However, the menu self-explanation tutor was completed significantly faster than the free-response tutor ($p = 0.029$).

4.1.3 Pilot: Discussion. The learning results show that self-explanation is effective for helping students learn to identify the correct article choice. However, unlike results from math and science domains, there was no advantage for self-explanation over a traditional practice environment. In fact, due to the increased amount of time needed for free-form explanations, self-explanation, especially in that form, may be an inefficient way to acquire the material. One of the limitations of the pilot study was that students in all conditions had at least some exposure to tutored practice (i.e., students in the two self-explanation conditions first selected the article to complete the sentence and then explained their choice). Thus, I cannot tease out the effects of self-explanation alone. In order to further investigate these issues and better evaluate the relative efficiencies of instructional conditions, I conducted a full-scale study exploring the effects of self-explanation alone versus practice alone. An additional result of this study was I gained trust and confidence with our partner instructors by demonstrating that the tutors met their educational objectives and all conditions led to significant learning gains. With that established, I was able to expand the number of participants in all subsequent studies.

4.2 Study 1: Self-Explanation vs. Practice

In Study 1, I compared two instructional conditions: tutored practice and self-explanation only. I chose to continue having students select explanations from a menu rather than generate their own in a free-form style for a variety of reasons. Perhaps most important, the pilot study showed that selecting explanations from menus was a more efficient strategy than free-form explanations. In addition, others have argued that feedback on explanations, a process facilitated by explaining with menus, can improve learning (Aleven & Koedinger, 2002). Finally, previous studies in STEM domains have shown explaining via menus is an effective approach (Aleven & Koedinger, 2002; Atkinson & Merrill, 2003).

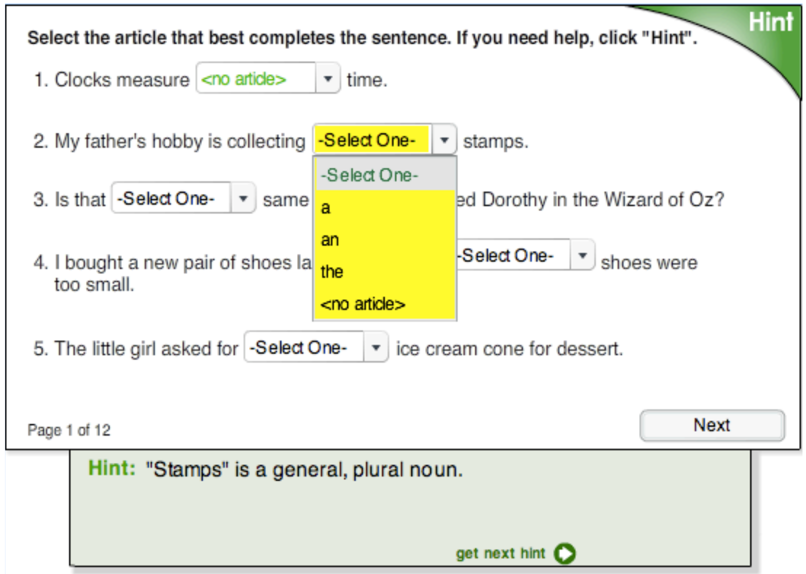


Figure 9. Tutored Practice. Using the practice tutor, students select the article that best completes the sentence. Students receive immediate feedback and have access to hints.

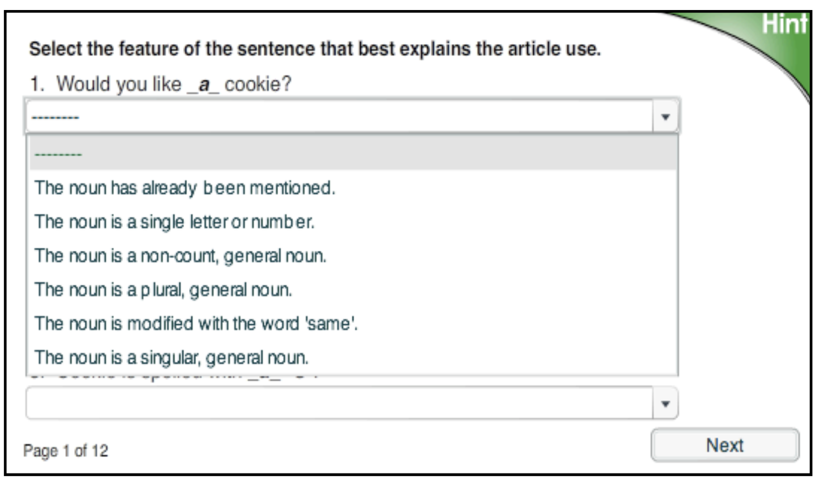


Figure 10. Self-Explanation Only. In this condition, students choose the rule or feature that best explains the article use for the given sentence.

4.2.1 Study 1: Conditions

The tutored practice condition was identical in form to the practice tutor used in the pilot study. However, I modified the self-explanation condition in order to understand if self-explanation alone, without article selection practice, is beneficial for learning and efficiency. While in the pilot study, students in the self-explanation conditions selected an article to complete the sentence before explaining, in this study, students were given a sentence with the correct article highlighted and only had to choose the rule or feature driving the article decision (Figure 10). The order that the rules were displayed in the menu was kept constant to reduce the search time students needed to select their answers. Again, students received immediate feedback on their answers and had access to hints.

In addition to the between-subjects practice versus self-explanation manipulation, I designed a within-subjects manipulation of tutored versus non-tutored knowledge components (as discussed in Chapter 3). Students were assessed on twelve knowledge components but only tutored on eight. The goals of this manipulation were to test whether the hypothesized knowledge components delimit the scope of transfer of learning (as KCs, by definition, should do) and to isolate the effects of computer-based tutoring as opposed to other factors (e.g., familiarity with the interface, fatigue, instruction outside the tutor etc.) that might effect the non-tutored items as well as the tutored ones. In order to include this manipulation, I added four KCs to the list of those used for the pilot study. Again, this was done in collaboration with classroom instructors and the added KCs were selected because they covered additional uses of "a", "an", and "no article". See appendix for complete list of KCs covered in this study.

4.2.2 Study 1: Measures

As in the pilot study, I assessed students on both article-selection and explanation-selection items. The article selection items were of the same form as those in the practice tutor, and the explanation selection items were of the same form as those in the self-explanation only tutor. Two isomorphic forms of each test type (article selection and explanation selection) were created and counterbalanced. This study design enabled assessment of both learning within each task type as well as transfer from one type to the other, that is, from practice to explaining and from explaining to practice.

4.2.3 Study 1: Procedure

The materials, random assignment and setting were the same as the pilot study. All students (N=117) were adult English language learners who participated as part of their normal grammar class (intermediate n=30, advanced intermediate n=61, advanced n=26). Since data collection was limited to a single class period, not all students completed the pretests, tutored instruction, and both posttests. Overall, 86% (101 out of 117) of the students completed all tasks; however, attrition between conditions was not the same with 95% (55 out of 58) of students in the tutored practice condition and 77% (46 out of 59) of students in the self-explanation condition completing all tasks (Fisher's exact test, $p = 0.01$). Despite the difference in attrition, the pretest scores for the target instruction (the article choice pretest) from students in both conditions who completed all tasks remained nearly identical ($F(1, 99) = 0.002$, $p = 0.96$) (Table 10). Based on observation and anecdotal comments from the classroom teachers, students who ran out of time before completing the posttests fell into one of two categories: students with low prior knowledge and students with high prior knowledge who were very meticulous and conscientious in their choices[1]. One student did not take the pretest due to a technical error and was dropped from analysis.

Table 10. Pretest scores before and after attrition were not different nor were there any differences between conditions before or after.

	Article Choice Pretest (Total Sample)		Article Choice Pretest (After Attrition)	
Tutoring Condition	n	Pretest Mean (SD)	n	Pretest Mean (SD)
Practice Tutor	58	65.3% (16.9)	55	65.7% (17.2)
Menu SE Tutor	59	65.7% (18.1)	46	66.0% (17.4)

[1] In Wylie, Koedinger, and Mitamura (2010) we report on a propensity analysis to adjust for possible differences and the results were no different from those reported below.

4.2.4 Study 1: Results

Learning Gains. Similar to the pilot study, repeated measures ANOVA analyses showed that students demonstrate significant pre to posttest learning gains on the article choice measure ($F(1, 99) = 39.9$, $p < 0.001$), but there was no interaction between condition and learning gain ($F(1, 99) = 0.072$, $p = 0.79$) (Table 5). Post-hoc analyses showed that the learning gains are significant for both conditions (practice: $F(1,54) = 20.4$, $p < 0.001$; explain: $F(1,45)=19.5$, $p < 0.001$). Similar analysis for the explanation (declarative knowledge) measure showed that, unlike in the pilot study, students showed significant pre to posttest gain ($F(1,99) = 13.2$, $p < 0.001$) (Table 12) and there is a marginal interaction between condition and gain ($F(1,99) = 2.76$, $p = 0.10$). Post-hoc analyses revealed that students in the explanation condition show a significant learning gain on the explanation measure ($F(1,45) = 10.3$, $p = 0.002$), and students in the practice condition show a marginal effect ($F(1,54) = 2.7$, $p = 0.10$).

Table 11. Article choice test scores, gain, and instructional time by condition. Students in both conditions show equal and significant gains, while students in the tutored practice condition (Figure 9) complete the tutor in significantly less time.

Condition	n	Article Pretest (SD)	Article Posttest (SD)	Average Normalized Gain (SD)	Instructional Time (SD)
Practice	55	67.3% (16.2)	80.0% (16.8)	38.3% (42.5)	11.2 mins (4.3)
Explanation	46	67.1% (17.4)	81.0% (15.5)	40.6% (44.3)	15.0 mins (5.0)
Averages	101	67.2% (16.7)	80.4% (16.1)	39.4% (43.1)	12.9 mins (5.0)

A hypothesized difference between the two conditions is that they facilitate the acquisition of different types of knowledge. The practice condition fosters procedural knowledge learning. Procedural knowledge is knowledge that is used when completing a task or knowing *how* to do something. In contrast, the explanation choice condition teaches declarative knowledge. In the case of the article system, declarative knowledge includes the grammar rules or knowledge components (cf., Anderson & Lebiere, 1998). These results showed that both groups are learning the skill on which they were tutored. Students in the practice condition show improvement on the procedural knowledge assessment (the article choice task), and students in the menu self-explanation group show improvement on the declarative knowledge assessment (the explanation choice task). More surprisingly, both groups also learned the transfer skill. Article choice students showed improvement on the explanation items, and explanation choice students showed improvement on the article choice items. Despite having no tutored practice on explanation choice items, students in the article choice condition improved in explanation choice. This procedural to declarative transfer, that is, improvement in verbal declarative knowledge as a consequence of procedural practice, was also observed in Aleven & Koedinger (2002) and may be a consequence of the verbal feedback and hint messages the tutor provides and perhaps some students spontaneously engaging in self-explanation without prompting. This procedural to declarative transfer may have been weaker than the more direct declarative to declarative transfer given the trend toward greater gains in explanation by the explanation condition than by the article-choice condition.

Table 12. Explanation Choice Test Scores, and Gain by Condition. Students in both conditions are showing significant learning gains on the explanation assessment, with students in the explanation-only group demonstrating marginally higher gains.

Condition	n	Explanation Pretest (SD)	Explanation Posttest (SD)	Average Explanation Normalized Gain (SD)
Tutored Practice	55	55.9% (28.2)	60.5% (23.0)	3.6% (20.6)
Menu SE-only	46	63.9% (23.5)	76.1% (22.8)	29.1% (25.9)
Averages	101	59.5% (26.3)	67.6% (24.1)	15.2% (23.4)

Similarly, I also observed declarative to procedural transfer. Students using the explanation choice tutor showed significant improvement on performance of the procedural task of making article choices. This effect can also be explained by Aleven & Koedinger's (2002) model that procedural tasks can be solved by *either* procedural *or* declarative knowledge. That declarative to procedural transfer was as strong as procedural to procedural transfer is consistent with this model.

I repeated the analysis for the untutored knowledge components with results showing no improvement for the article choice items (pretest mean = 76.5% (SD=21.5), posttest mean = 73.5% (SD=26.4) $F(1, 99) = 1.13$, $p = 0.290$) and a significant decrease on the explanation choice items (pretest mean = 54.0%, posttest mean = 48.5% $F(1, 99) = 5.19$, $p = 0.025$). These results support the hypothesis that our proposed knowledge component model reasonably explains the scope of transfer. A knowledge component acquired during tutoring (whether in declarative/explicit or procedural/implicit form) transfers only to tasks associated with it and not to others. These results also indicate that the observed learning gains on the tutored rules were the result of students' experiences with the tutors and not an unobserved, external factor. The decrease in explanation performance on untutored items may be a result of students' tendency to select explanations consistent with the tutored items, which, of course, are incorrect for the untutored items. To test for this possibility I compared the number of times a tutored explanation was chosen for a non-tutored knowledge component on both the pretest and posttest and indeed found a significant difference. Students were more likely to choose a tutored explanation for a non-tutored item on the posttest than they were on the pretest ($t(100)=2.46$, $p = 0.02$), suggesting that students developed a leaning towards selecting tutored explanations.

Instructional Efficiency. Data on instructional time, that is, the time spent using the tutors, revealed a significant difference between the two conditions. Students using the self-explanation tutor spent almost 50% more time using the system (M=15.0 minutes, SD=5.0) than students using the practice tutor (M=11.2 minutes, SD = 4.3 $F(1,99)=16.6$, $p < 0.001$). When considering instructional efficiency together with the learning gain results, it is important to remember that the target skill for this domain is article selection not explanation. Since students in both conditions demonstrated equal article selection gain, the timing results suggest that even though self-explanation helps students learn, in this domain, it is less efficient than tutored practice.

4.2.5 Study 1: Discussion

In many ways, these results replicate what was observed in the pilot study, mainly that prompting students to self-explain while learning the English article system has trade-offs. First, these results again demonstrate that the tutors themselves are effective. Students demonstrated significant learning gains on the tutored but not the untutored knowledge components. Next, results show that self-explanation alone, even when not paired with practice, does yield learning and transfer to procedural tasks. However, there is no difference on the more pedagogically relevant article selection task, and students spend significantly more time working with the self-explanation tutor suggesting that self-explanation is an inefficient strategy in this domain.

Results showed declarative to procedural transfer that is as strong as procedural to procedural transfer, though it takes substantially longer for such learning to take hold. I also found procedural to declarative transfer, although it is weaker than declarative to declarative transfer. While this declarative knowledge is arguably of no practical importance in this domain, or similar domains where procedural fluency is the prime objective, this result is of theoretical interest and I return to it in the general discussion.

One reason that the self-explanation effect may not generalize to language learning could be due to the metalinguistic challenges and increased cognitive load that students face when doing self-explanations in their non-native languages. For example, many of the article rules contain challenging and domain-specific vocabulary that may be difficult for a non-native speaker. For instance, the words “singular” and “consonant” in "Use 'a' when the noun is general, singular, and begins with a consonant sound" are relatively infrequent words with specialized meanings in the language context. Perhaps a better strategy would be one that continues to foster deep processing like self-explanation but decreases the metalinguistic demand. Thus, these results prompted further exploration into the instructional design space in an attempt to find an instructional strategy that might reduce extraneous cognitive load (Sweller, 1988) by decreasing metalinguistic demands.

4.3 Study 2: Analogical Comparisons, Self-Explanation, and Practice

To follow-up with Study 1, which examined explanation-only and tutored practice conditions, I conducted a follow-up study that employed a new instructional condition. In Study 2, I developed a new version of the tutor intended to reduce metalinguistic difficulty compared to the self-explanation condition. To compete against self-explanation, I chose to implement a modified version of analogical comparison (Gentner, et al., 2009; Gick and Holyoak, 1983). In a typical analogical comparison problem, students are presented with two worked examples and asked to compare the similarities and differences between them. The theory behind analogical comparisons is that by comparing the examples, students will be able to induce the underlying schema of the two problems (Gentner et al., 2009; Gick & Holyoak, 1983). Like self-explanation, analogy training has proven to be successful for a variety of domains and learners. In a study investigating business negotiation training, Gentner, et al. (2003) found students who were instructed using analogical encoding produced better written solutions on posttest items and were able to transfer their skills to the more challenging modality of face-to-face negotiation. In Study 2, I tested the hypothesis that analogical comparisons will lead to greater learning and increased efficiency compared to prompted self-explanation because analogical comparisons should reduce the metalinguistic demands imposed on the student. As seen in the self-explanation tutor in Figure 10, many of the explanations have domain-specific vocabulary words (e.g., The noun is *singular*, *general*, and begins with a *consonant* sound). In this explanation, students must tackle vocabulary words (e.g., "consonant") and phrases (e.g., "general noun") whose specialized meanings may be unfamiliar or at least are not well practiced.

To reduce the linguistic demands compared to the self-explanation tutor, I designed a tutor version where students select an analogous sentence that used more familiar words, instead of an explanation that often used highly specialized metalinguistic terms. The process of selecting an analog sentence may support students in deeply processing the material, making mappings across analogs to encourage identifying the deep features important for making article decisions. This analogical induction process can be performed without the metalinguistic demands of explanation processing. Furthermore, the analogous sentences used for the comparison provide the added advantage of presenting students with more examples of correct article use, which alone may be beneficial for language learning even without simultaneous comparison (Mitchell & Myles, 2004).

I also made three other design changes. As in the pilot study, the self-explanation instruction condition contained both practice opportunities *and* prompts to self-explain, however in a slightly different format. In the pilot study, students were asked to both practice (select the correct article) and explain (select or type an explanation) for the same sentence. In Study 2, I separated these tasks such that for a given sentence, students *either* chose the article *or* chose the explanation. Previous work has suggested a benefit for explaining expert solutions instead of a student's own work (Calin-Jagerman & Ratner, 2005). I also interleaved the self-explanation and practice problems. This integration of different problem types is common in past studies (e.g., Renkl, et al., 2002; Rau, et al., 2009) and is motivated by the theory that there are complementary benefits of explaining and practicing (cf. Matsuda, et al., 2008). The other change I made was to remove the explanation assessment in order to reduce the overall time required to complete the study and avoid issues surrounding attrition. These changes are described in more detail below. Finally, in order to increase the number of opportunities students had for each particular knowledge component, I reduced the number of KCs to six. The six KCs covered two rules for each of the three article choices: a/an, the, and no article. A complete list of the KCs used in this study can be seen in the appendix.

4.3.1 Study 2: Problem Types

This study employed three types of tutored problems: practice, menu self-explanation, and analogical comparison. The practice problems were again identical in form to those used in the previous studies: students were given a sentence and chose the article (a, an, the, or no article) that best completed the sentence. The self-explanation problems were similar to those from Study 1. Students were given a target sentence with the article highlighted and chose the explanation from the provided menu. The analogical comparison problems were similar to the self-explanation problems except instead of choosing the explanation, students chose the analogous sentence that used the same article rule as the given sentence (Figure 11). For example, given the sentence pair "*Last week, I bought a car. Today, the car broke*." students should choose the sentence "*Sally found a dog, and the dog is small and black,"* since both the given and analogous sentences use the rule: *If a noun has* *already been mentioned then use "the"*. There was one analogous sentence for each of the six article rules covered in the material. In an attempt to prevent students from developing spurious associations, all the analogous sentences were approximately equal in length and used similar vocabulary. In addition, the analogous sentences used simple vocabulary and were easy to read (Flesch-Kincaid Grade Level = 2.0). The same six analogous sentences were presented in the same order for each of the analogy problems.

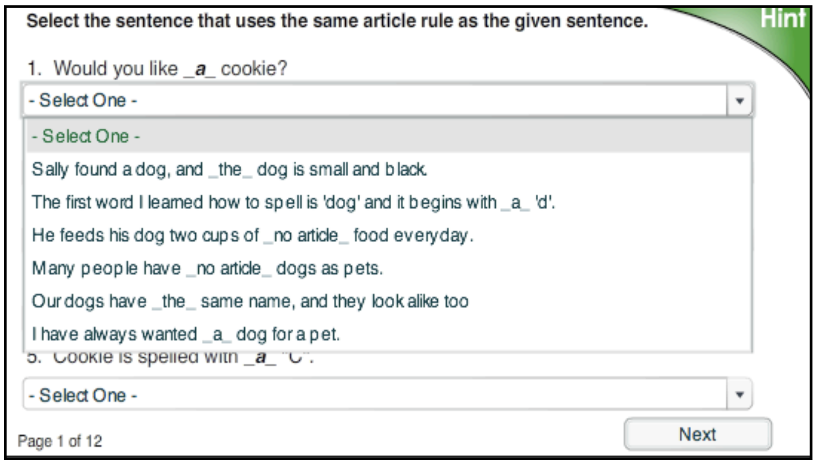


Figure 11. Analogy. In the analogy problems, students select the example sentence that uses the same article rule as the given sentence.

4.3.2 Study 2: Conditions

Three corresponding experimental conditions were created using the above task types: tutored practice, self-explanation with practice, and analogy with practice. Students in all conditions received 30 identical practice problems. In addition, students in each condition received 30 condition-dependent items: students in the self-explanation with practice condition received 30 self-explanation problems, students in the analogy with practice condition received 30 analogy problems, and students in the tutored practice condition received 30 additional practice problems.

Previous research has shown the benefits of interleaving examples with problem-solving practice (Pashler, et al., 2007; Trafton & Reiser, 1993) and that learning from examples is more beneficial during early rather than later stages of skill acquisition (Atkinson, et al., 2000). Therefore, I had students in the self-explanation with practice and analogy with practice conditions do more condition dependent items in the beginning and then move to interleaved blocks of matched practice and condition dependent problems, and finally, end with practice problems. More specifically, in the self-explanation with practice condition, the first ten problems were self-explanation problems; the next forty problems consisted of alternating blocks of five practice problems and five explanation problems, and finally, students completed ten practice problems. The analogy with practice condition used the same structure but students did analogy problems instead of self-explanation items (Table 13).

Table 13. Breakdown of problem type by condition. Students in the tutored practice condition completed sixty practice problems. Students in the Analogy with Practice and Explanation with Practice conditions saw analogy or explanation items interleaved with practice.

Item #	Tutored Practice	Analogy with Practice	Explanation with Practice
1-5	Practice	Analogy	Explanation
6-10	Practice	Analogy	Explanation
11-15	Practice	Practice	Practice
16-20	Practice	Analogy	Explanation
21-25	Practice	Practice	Practice
26-30	Practice	Analogy	Explanation
31-35	Practice	Practice	Practice
36-40	Practice	Analogy	Explanation
41-45	Practice	Practice	Practice
46-50	Practice	Analogy	Explanation
51-55	Practice	Practice	Practice
56-60	Practice	Practice	Practice

I controlled for several factors in the design of the three conditions: all conditions used the same sixty target sentences, presented in the same order, and the hints presented the same information, although in slightly different forms. For the practice problems, the first hint presented the features of the sentence relevant to article choice (e.g., The noun "plums" has already been mentioned.); next, students were given the complete rule, and finally, students were told which article to select. When completing the explanation problems, students were first presented with the relevant features of the sentence, and then told which explanation to choose. For the analogy problems, students first saw the relevant features; second, they were given the example sentence that contained the same feature; and, finally, told which example sentence to select.

4.3.3 Study 2: Procedure

Again, data collection for this study was conducted as part of normal classroom practice with students (N=99) from the English Language Institute at the University of Pittsburgh. Genders were equally represented, and students were enrolled in one of three course levels (intermediate n=24, advanced intermediate n=41, advanced n=34). After a brief introduction to the tutoring systems, students completed a computer-based pretest and were randomly assigned to one of three conditions: tutored practice, analogy with practice, or self-explanation with practice. Immediately after finishing the tutoring, students completed the posttest, which was isomorphic and counterbalanced with the pretest. In an attempt to have all students complete the instruction and assessments, in this study, I only used the target items (article selection tasks) for the pre and posttest. As in the previous studies, students chose the best article to complete the sentence and did not have access to hints or receive feedback on their selections. By removing the explanation selection assessment, all students completed all tasks.

4.3.4 Study 2: Results

Learning gains. Results of a repeated measures analysis of variance (ANOVA) with test score as the dependent measure, test time (pretest and posttest) as a within-subject factor, and tutoring condition as a between-subject factor revealed a significant main effect for test time ($F(1,96) = 63.8$, $p < 0.001$) and no interaction of test time by condition ($F(2, 96)=1.42$, $p = 0.25$). Again, students, regardless of condition, demonstrated significant learning gains (Table 14).

Table 14. Pretest, Posttest, Gain, and Instruction Time by Condition. All conditions demonstrated significant pre- to posttest learning gains and gains by condition were statistically indistinguishable. However, students using the practice tutor completed the instruction significantly faster than those using either of the deep-processing (explanation with practice or analogy with practice) conditions.

Condition	n	Pretest (SD)	Posttest (SD)	Average Normalized Gain (SD)	Instructional Time (SD)
Tutored Practice	33	60.1% (23.0)	77.5% (15.6)	39.5% (30.7)	13.4 mins (4.3)
Explanation with Practice	32	66.4% (16.7)	80.5% (9.3)	35.6% (29.7)	18.6 mins (6.0)
Analogy with Practice	34	66.2% (15.3)	76.5% (16.9)	29.8% (38.6)	17.0 mins (7.5)
Total	99	64.2% (18.7)	78.1% (14.4)	34.9% (33.3)	16.3 mins (6.4)

Instructional Efficiency. An ANOVA with total time spent using the tutor as the dependent variable reveals a significant effect of condition ($F(2, 96) = 6.45$, $p = 0.002$). Post-hoc Tukey HSD tests revealed that students in the tutored practice condition completed the instruction the fastest (M=13.4 minutes, SD=4.3), significantly faster than those in the analogy with practice condition ($p=0.04$, M=17.0 minutes, SD=7.5) and the self-explanation with practice condition ($p=0.002$, M=18.6, SD=6.0). No significant difference was found between the time-on-task of students in the two deep-processing conditions, self-explanation with practice and analogy with practice ($p=0.51$).

4.3.5 Study 2: Discussion

The results of Study 2 reveal again that regardless of condition, students demonstrated significant pre to posttest learning gains, and that students in the deep processing conditions (analogy with practice and self-explanation with practice) spend significantly more time using the tutor than those in the tutored practice condition. These results replicated the findings from Study 1 and show that the practice tutor is the most efficient strategy for this domain. Again, these results can be explained by the type of knowledge that students gain working with each of the conditions. Perhaps both the self-explanation and analogy problems facilitated declarative but not procedural learning, and while this knowledge is beneficial in that it can later be used to solve the procedural (article selection) tasks, it is not as efficient.

Again, the learning process evoked by the tutored practice condition appears to more efficiently produce procedural skill of equivalent accuracy than does self-explanation or analogical comparison. The hypothesis that students using the analogy with practice tutor would require less time than students using the explanation with practice tutor was not confirmed.

4.4 Study 3: Example Study, Explanation, and Practice

A consistent theme has emerged from the previous studies: self-explanation is an effective strategy but it takes significantly more time to complete with no resulting additional benefit. These results highlight the importance of investigating instructional efficiency, and in study 3, I investigate this issue further by taking a simpler approach and comparing efficiency results with the practice tutor. For this study, I developed an example study tutor in which I removed the explicit self-explanation prompts and simply asked students to study the correct examples provided. The idea of asking students to study correct examples is based on the success of using worked examples as an instructional strategy. In a worked example problem, instead of solving the problem on their own, students are given the answer and study the procedure by which the answer was derived. Worked examples have shown to reduce instructional time (Clark and Mayer, 2004) while continuing to lead to robust learning results (Sweller & Cooper, 1985; Paas & van Merrienboer, 1994; Atkinson, et al., 2000). Since English articles require only one step to complete (students implicitly or explicitly identify the features and then select the appropriate article), the example study problems simply present the student with the correct article highlighted and prompt the student to study the provided example.

However, just as the previous work on self-explanation has been done primarily in science and technology domains so has the majority of the work investigating the effects of studying worked examples. Perhaps the time savings seen in the previous studies will not replicate when applied to second language learning. One hypothesis is that the patterns of relevant features are more apparent in math domains than in language learning and thus can be more easily extracted during worked example study.

4.4.1 Study 3: Conditions

I designed Study 3 both as a replication of Study 2 to gain further evidence that there are no learning differences between self-explanation and practice and to examine the effects of the example study condition. The design of this study was nearly identical to Study 2, but the analogy with practice condition was replaced with an example study with practice condition. In the example study problems (Figure 12), students were given a sentence with the correct article highlighted (a worked example) and asked to study the example and to check the corresponding checkbox when they were finished. No other action by the student was required - they neither had to select the article nor explanation; however, they had access to the same hint sequence if they would like more information and an explanation for the given article use.

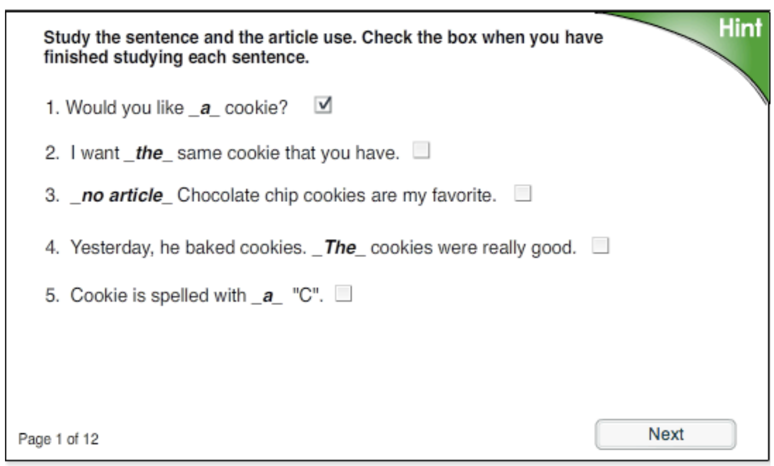


Figure 12. Example Study. In this condition, students read and study the given sentences, checking the box to indicate that they are finished.

4.4.2 Study 3: Procedure

Again, as mentioned above, students (N=96) participated as part of normal classroom practice and were enrolled in either the intermediate (n=16), advanced-intermediate (n=45) or advanced (n=35) grammar course at the University of Pittsburgh's English Language Institute. Data collection was completed in one class period. After a brief introduction on how to access and use the tutors, students took a pretest consisting of article choice items. They were then randomly assigned to a tutoring condition (practice, explanation with practice, or example study with practice). Finally, students took an immediate posttest, which was isomorphic to and counterbalanced with the pretest.

4.4.3 Study 3: Results

Learning gains and instructional efficiency. Similar to the previous studies, repeated measures analysis of variance results showed that students show significant learning gains ($F(1, 90) = 51.1$, $p < 0.001$) but with no effect of condition ($F(2, 90) = 0.48$, $p = 0.62$). And again, there was a significant effect of condition for time spent using the tutor ($F(2,90) = 3.941$, $p = 0.02$) (Table 15).

Table 15. Pretest, Posttest, Gain, and Instruction Time by Condition. Across all conditions, students show significant pre- to posttest improvement. Students in the tutored practice condition complete the instruction significantly faster than those in the explanation with practice condition.

Condition	n	Pretest (SD)	Posttest (SD)	Average Normalized Gain (SD)	Instructional Time (SD)
Tutored Practice	33	71.7% (15.6)	83.1% (11.1)	37.2% (35.9)	13.9 mins (4.6)
Explanation with Practice	30	62.8% (19.0)	76.9% (15.4)	37.6% (37.9)	17.4 mins (5.4)
Example Study with Practice	33	72.7% (12.5)	83.4% (12.9)	39.7% (35.6)	14.4 mins (5.4)
Average	96	69.3% (16.4)	81.3% (13.4)	38.2% (36.0)	15.2 mins (5.3)

Post-hoc Tukey HSD tests revealed that students in the practice tutor complete the instruction the fastest (M=13.9 minutes, SD=4.6) and significantly faster than those in the self-explanation with practice condition (M=17.4 minutes, SD =5.5 p = 0.03). In addition, the example study with practice condition was marginally faster than the self-explanation with practice condition (M = 14.4 minutes, SD = 5.4, p =0.06), but there was no difference between the tutored practice and example study with practice conditions (p=0.93).

A look at the usage of hints between conditions suggests why students in the example study condition did not complete the instruction as fast as hypothesized. There was a main effect for condition on hint use (F(2,93)=9.00, p < 0.001), with post-hoc Tukey HSD showing that students in the example study condition asked for significantly more hints than those in either the practice (p = 0.001) or self-explanation conditions (p = 0.001). While students in the practice and self-explanation condition asked for a hint on less than 10% of the 60 tutor questions (practice condition mean hint use = 4.82 questions (SD = 8.31), self-explanation condition mean hint use = 4.80 questions (SD = 6.24)), students in the example study condition asked for a hint on approximately 25% of the items (example study mean = 15.42 (SD = 17.02)).

4.4.4 Study 3: Discussion

The general pattern of results from Study 3 is similar to those found in the earlier studies. Students in all conditions demonstrate significant learning gains and students in the practice condition completed the instruction significantly faster than those in the self-explanation condition. While I had thought that the example study condition would be completed the fastest since the worked example problems theoretically demand less cognitive load than the other problem types, timing data show no significant difference between completion times for the example study with practice and the tutored practice condition and a marginal effect for the self-explanation with practice condition.

It is clear from the hint use result that students do have a desire to make use of verbal declarative rules during learning. Whether or not it aids their learning is unclear. Perhaps reading hint messages facilitates the learning process. However, they didn't learn more. So, the alternative is that the example study condition is encouraging students to make a bad metacognitive choice and ask for hints even when they are not beneficial, and the extra hint usage is burning time and not contributing to learning. Consistent with this second alternative, I find no correlation between hint use and learning for students in the example condition.

4.5 Study 4: Adaptive Self-Explanation

One limitation of these previous findings is the lack of robust learning measures that test long-term retention or declarative knowledge acquisition. Robust learning is more likely to become apparent on delayed rather than immediate posttests (Craik & Kockhard, 1972), and previous work has found benefits for self-explanation on robust learning measures even when no immediate benefit was apparent. When teaching middle school students science inquiry strategies for designing non-confounded experiments, Sao Pedro and colleagues (2010) compared direct instruction with and without self-explanation prompts. Their results showed no difference between conditions on the immediate posttest, but there was a significant advantage for the self-explanation group on a delayed posttest six months after training. In addition, in work investigating the effects of explicit metalinguistic feedback compared to implicit recast feedback, Ellis, et al., (2006) found that while there were no difference between groups on the immediate posttest, the metalinguistic group outperformed the recast group on the delayed (12 days post instruction) posttest on several different measures. In addition, others have found that self-explanation helps students acquire different types of knowledge than practice alone. In their work in high school geometry, Aleven and Koedinger (2002) found that students in a problem-solving (no self-explanation) condition gained more *procedural* knowledge than students prompted to self-explain, resulting in better performance on simple numeric problems. However, students who were prompted to self-explain developed better *declarative* knowledge and thus demonstrated greater accuracy on the more challenging transfer items. One theory hypothesizes that self-explanation provides an additional pathway (through the use of declarative knowledge) by which students can solve problems (Chi, et al., 1994). This declarative knowledge pathway may be more generalizable and less susceptible to decay than procedural knowledge alone.

The primary objective of the final study was to investigate whether the effects of self-explanation on long-term retention and knowledge acquisition (both procedural and declarative) generalize to the kinds of knowledge needed for second language grammar learning. In this section, I describe the design of a new adaptive self-explanation tutor, in which students are prompted to explain the article use in a sentence only if their first attempt to selecting the article is incorrect, and then present the results of the tutor compared to a practice-only (no self-explanation) tutor in a classroom study with adult English language learners.

4.5.1 Study 4: Power Analysis

Given that the previous studies had resulted in no significant differences between the self-explanation and tutored practice conditions on procedural knowledge measures, it was especially important to calculate the estimated statistical power. Power refers to the probability that the null hypothesis will be rejected when it is false. Using data from the previous studies and conservatively estimating 35 students per condition, I calculated power[2] for condition differences of 10% (approximately one letter grade) as well as 20% (approximately two letter grades), with results of 0.73 and 0.92 respectively (Figure 13). These results suggest that the study has high power to detect a difference of two letter grades between the two conditions and medium power to detect the smaller difference of one letter grade between the two conditions.

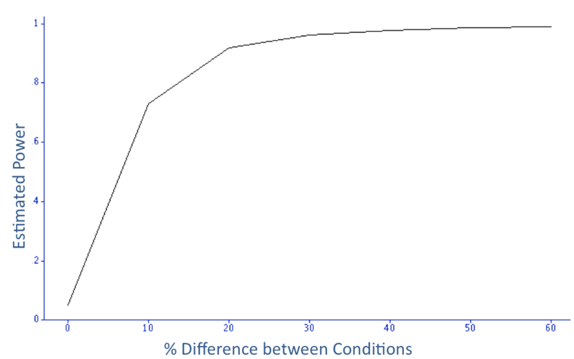
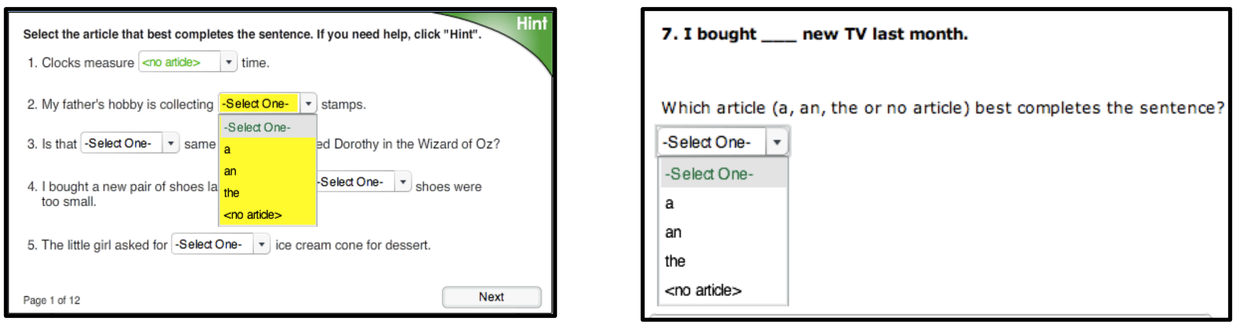


Figure 13. Estimated power for differences between conditions ranging from 10-60%.

[2] Power calculated using the power analysis tools found at: http://www.cs.uiowa.edu/~rlenth/Power/

4.5.2 Study 4: Conditions

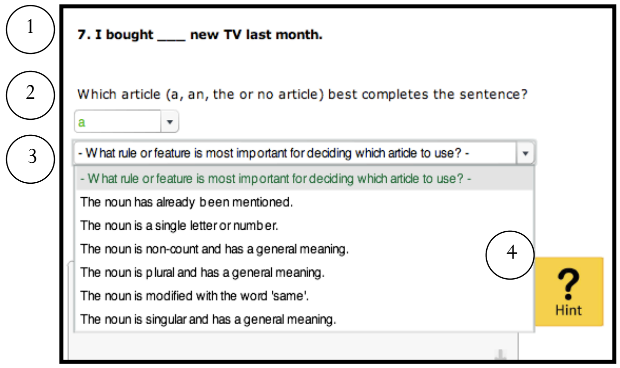
For this study, the practice tutor was slightly modified in order to make it more similar in appearance to the adaptive condition. However, the student behaviors, feedback, and hints remained unchanged from the previous studies. The only difference was that instead of seeing a screen with multiple sentences, students saw one sentence at a time (Figure 14).



Previous Studies Current Study.

Figure 14. For the final study, the tutored practice condition changed slightly in that sentences were presented one at a time rather than in groups.

In the previous studies, self-explanation proved to be an effective but inefficient strategy. However, in these earlier studies, the self-explanation prompts were static; all students using the self-explanation tutor were prompted to explain the same number of sentences. However, work by McNamara and colleagues suggests that prompted self-explanation is especially beneficial for students with low prior knowledge (McNamara & Scott, 1999). Thus, in an attempt to make a more efficient tutor, I built an adaptive self-explanation tutor that only prompts students to self-explain when estimates of their prior knowledge for a given article rule are low. The method for deciding when to prompt students to self-explain is similar to the method the ASSISTments system uses to provide additional scaffolding (Razzaq, et al., 2005). Namely, students are first presented with a cloze (sentence completion) task. If they choose the correct article on the first attempt, they move on to the next sentence. If students make an error or ask for a hint on their first step, I posit that they do not know the rule. Thus, after eventually selecting the correct article, they are prompted to self-explain by choosing the appropriate explanation from the provided menu (see Figure 15).



① Students are presented with a single cloze sentence at a time.

② Students select the article that completes the sentence.

③ If students do not select the correct article on their first attempt, they continue making article choices until correct and are then prompted to self-explain.

④ Students receive immediate feedback on each selection and have access to hints for article selection and explanation steps.

Figure 15. Adaptive self-explanation tutor. Students are only prompted to self-explain if their first attempt at selecting the article is incorrect. The practice tutor is identical except students are never prompted to self-explain.

4.5.3 Study 4: Participants

Participants were adult English language learners (M=25.5 years, SD=5.3) enrolled in one of three levels of an intensive language program: intermediate, intermediate-advanced, advanced. Instruction and assessments were incorporated into normal classroom activities. As in the previous studies, genders were equally represented, and students spoke a variety of first languages. In total, I collected tutor data from 131 students. However, due to attrition, I report on data from the 92 students who completed all tasks (tutoring and all four article-selection assessments). Of the 39 students who were dropped, there was no difference in attrition between conditions $\chi^2(1, N=131) = 0.084$, $p=0.77$) and no difference in immediate learning gains between those who were dropped (M=12.5%, SD=13.3) and those included in the analyses (M=7.9%, SD=18.1, $F(1, 124)=1.84$, $p=0.18$).

4.5.4 Study 4: Measures

In this study, the primary measures for learning consisted of a procedural knowledge (article selection) and a declarative knowledge (article feature) assessment. The procedural knowledge assessment was of the same form as that used in the previous studies and consisted of problems similar to the cloze tasks that students completed as part of the tutoring (e.g., *Yesterday, I bought a new car. ____ car is red.*). For each sentence, students chose the best article (*a*, *an*, *the*, or *no article*) to complete the sentence. Yet, unlike during tutoring, students did not receive feedback on their answers and did not have access to hints during the tests. The procedural knowledge assessment was administered at four different points in time: pretest, immediate posttest, one-week after tutoring, and approximately two months after tutoring. Four isomorphic versions of the assessment were created; two were counterbalanced as the pretest and posttest, one was the one-week retention for all students, and one was the two-month retention for all students. In order to have measures of comparable difficulty, the two long-term retention measures consisted of the most difficult items from the tutor.

Results in other domains suggest that self-explanation works by increasing a student's declarative knowledge, and declarative knowledge might transfer to better procedural ability. To test whether this correlation between declarative and procedural knowledge exists in language learning, I included a declarative knowledge assessment that presented students with a relevant article feature and asked them to choose which article would be used (e.g. *If a noun has already been mentioned, which article do you use?*). This assessment was given before and immediately after tutoring. The assessment was identical both times because it covered all six rules included in the tutoring and the nature of the question (asking students to map the unique feature to its specific responses) prevents creating isomorphic items. In addition to knowledge gain, I was also interested in learning efficiency since these interventions are designed for use in real classrooms. To determine efficiency, I computed the total time students spent working with tutor.

4.5.5 Study 4: Procedure

All tasks were completed as part of normal classroom practice. On the day of instruction, students met in the computer lab and began by taking the declarative knowledge pretest on paper. Students were then given a brief introduction to the tutoring systems and the article rules that would be covered (approximate length 5 minutes). Students next used the computers to complete the procedural knowledge pretest and were randomly assigned to a tutoring condition. After using the tutor, students used the computer to take both immediate posttests (procedural knowledge followed by declarative knowledge) as well as complete a demographic survey. Within two days of completing the instruction, students received a handout containing the sentences and their individual answers on the pretest and immediate posttest. This handout was provided for ecological validity reasons (it is custom for students to receive feedback on in-class assignments) and briefly reviewed during class. During the week between tutored instruction and the first retention test, students in the intermediate and advanced grammar courses received additional instruction on the article domain in general (the article tutor only covers a subset of all article rules) as part of normal classroom instruction. However, articles are not explicitly covered in the intermediate-adv course and thus the only follow-up instruction those students received was the handout. The second retention assessment was administered as part of the students' final exam.

4.5.6 Study 4: Results

An initial investigation of the adaptive condition shows that on average students using the adaptive tutor were prompted to explain 14.2 (SD=6.5) of the 60 sentences. As expected, there is a significant negative correlation between pretest score and number of explanation prompts ($r(45) = -0.30$, $p = 0.04$). This result suggests that the adaptive self-explanation tutor succeeded in targeting students with low prior knowledge.

When examining the learning and efficiency results, I first investigated the hypotheses generated from the previous studies: (H1) The practice and adaptive self-explanation tutors will lead to equal procedural learning gains, and (H2) the practice tutor will take less time to complete than the adaptive self-explanation tutor. I then examined the effects of self-explanation on declarative knowledge and long-term retention through the following hypotheses: (H3) Self-explanation will lead to more declarative knowledge than tutored practice, and (H4) Self-explanation will lead to better long-term retention than tutored practice.

Results replicated the findings of the previous studies and confirm the hypothesis that both conditions lead to equal learning gains (H1). A repeated measures ANOVA on the procedural knowledge (article selection) assessment using the pretest and immediate posttest shows that students in both conditions demonstrate significant pretest to posttest learning gains ($F(1,88)=13.1$, $p=0.001$, $\eta^2=0.13$), but there was no difference between conditions ($F(1,88)=0.30$, $p = 0.58$) (Table 16) or course level (intermediate, intermediate-advanced, advanced) ($F(1,88)=1.66$, $p=0.20$). Efficiency results also replicated previous findings in this domain and confirm the hypothesis that the practice tutor was faster to complete than the adaptive self-explanation tutor (H2); students in the practice tutor completed the instruction significantly faster (M=15.0 minutes, SD=4.9) than students using the adaptive tutor (M=17.7 minutes, SD=4.3, $F(1,90)=7.8$, $p = 0.006$, $\eta^2=0.08$).

To test whether the adaptive self-explanation condition lead to more declarative knowledge gain than the practice tutor (H3), I conducted a repeated measures ANOVA on the declarative knowledge assessment. Again, both conditions led to significant pretest to posttest improvement ($F(1,77)=86.2$, $p<0.001$, $\eta^2=0.53$). Furthermore, results confirm hypothesis H3: The adaptive self-explanation tutor leads to greater declarative knowledge gains than the practice tutor ($F(1,77)=4.39$, $p=0.04$, $\eta^2=0.05$) (Table 16).

Table 16. Learning gains by condition for the procedural (article selection) and declarative (article feature) assessments. Students in both conditions show significant learning on both assessments, and students using the adaptive tutor make greater gains on the declarative assessment than students using the practice tutor.

	Article Pretest (SD)	Article Posttest (SD)	One-week Retention (SD)	Two-month Retention (SD)	Declarative Pretest (SD)	Declarative Posttest (SD)
Adaptive SE n=47	68.8% (14.2))	78.0% (12.9)	82.6% (14.8)	81.8% (17.5)	55.8% (22.5)	90.3% (14.3)
Tutored Practice n=45	71.3% (16.2)	77.8% (16.3)	86.2% (11.0)	83.8% (14.8)	62.8% (22.1)	84.6% (26.6)

For testing the final hypothesis (H4), I did a repeated measures ANOVA using all four instances of the procedural knowledge assessment. Again, overall, students showed consistent improvement with each test ($F(3,86)=24.8$, $p<0.001$, $\eta^2=0.46$) and there was an interaction between course level and test time ($F(3,86)=2.78$, $p=0.01$, $\eta^2=0.09$) but I find no evidence that self-explanation is better for long-term retention than practice-alone ($F(3,86)=0.56$, $p=0.64$) and thus cannot confirm H4 that self-explanation leads to better long-term retention. Post-hoc analyses showed that when controlling for pretest score, regardless of condition, students in the intermediate-advanced course (who did not receive post-tutor in-class instruction) did as well as students in the intermediate and advanced courses on the one-week retention test ($F(2,88)=1.44$, $p=0.24$) and marginally worse on the two-month retention test ($F(2,88)=2.65$, $p=0.08$). This is likely attributed to curriculum differences and the fact that students in the intermediate and advanced courses studied for the two-month retention exam while students in intermediate-advanced course did not.

In addition to the above hypotheses, the collected data allowed for various relationships to be investigated. For example, since some students were absent the day that tutoring occurred, I could look at correlations between whether a student used the tutor and scores on the one-week retention exam. Of the 127 students who took the one-week retention test, 116 were also present the day of tutoring and 11 were not. An ANOVA using one-week retention score as the dependent measure and attendance as the independent measure showed a significant effect for attendance. Students who were present the day the tutors were used did significantly better than those who were absent ($F(1, 127)=17.2$, $p<0.001$) (Table 17).

Table 17. Students who were present during tutoring did significantly better on the one-week retention than students who were absent.

	Average One-Week Retention Score (SD)
Students Present for Tutoring (n=116)	84.2% (13.8)
Students Absent for Tutoring (n=11)	65.7% (18.2)

4.5.7 Study 4: Discussion

One of the primary goals of the learning sciences is to understand when and why instructional manipulations succeed with the ultimate success being the discovery of a strategy that works for all domains. While self-explanation has previously been called a "domain general" strategy, these results suggest that there may be limits to its generalizability depending on the goals of instruction and the nature of the targeted knowledge. Specifically, the results of this study show that self-explanation is generalizable in that it leads to an increase in declarative knowledge over a comparable tutored practice condition. However, this additional knowledge does not transfer to better procedural performance, which, for this domain, is the primary goal of instruction. English language learners want to speak and write more accurately and generally do not need to explain their article choices.

These results contribute to developing a predictive theory of when self-explanation will help. The findings also replicate the previous studies and show that self-explanation is inefficient for this domain. Prompting students to self-explain takes longer than practice-alone and does not result in greater procedural learning. In the previous studies, I looked at several variations of this study all with similar results. Deep-processing manipulations, like self-explanation and analogical comparisons, are effective but take significantly more time and result in no additional pedagogically relevant benefit compared to manipulations designed to increase the rate of processing (tutored practice and example study). Further, these results show that even when more robust learning measures like long-term retention are considered, there is no evidence for an advantage of self-explanation over practice. However, I find that self-explanation leads to greater declarative knowledge acquisition across different types of domains, and thus for domains in which declarative knowledge is a learning objective or those in which declarative knowledge transfers to greater procedural ability, self-explanation should be both an effective and efficient strategy.

4.6 Summary Results

One clear finding is that self-explanation is effective for teaching students the English article system. Across all studies, students in the self-explanation conditions demonstrate significant pre to posttest learning gains. However, it is important to note that while self-explanation is effective, there is no evidence that self-explanation is more beneficial than tutored practice. In addition, for the studies in which I controlled for the number of items (Study 1, 2, 3, and 4), it is clear that the practice condition requires significantly less time to complete (Table 18). More broadly, these results provide evidence that boundary conditions exist for the effectiveness of self-explanation and that in the case of English article learning, prompting students to self-explain is an inefficient strategy. While interpreting null results is always challenging, the highly consistent results across all studies suggest that there really is no difference in effectiveness between practice with feedback and prompted self-explanation for acquiring the procedural knowledge needed to correctly select articles to complete a sentence.

However, the results suggest that there is a difference in learning gains with respect to declarative knowledge. In both Study 1(Self-explanation Only vs. Practice Only) and Study 4 (Adaptive Self-Explanation vs. Practice), students in the self-explanation conditions demonstrated greater gains than students in the tutored practice conditions. These results suggest that self-explanation still "works" in second language grammar learning in that it helps students acquire declarative knowledge; however, in this domain (unlike in math and science), the additional declarative knowledge does not aid procedural ability.

Table 18. Summary of learning gains. Results show that all instructional conditions are effective at teaching students the English article system, and there are no differences between conditions.

Study	Condition	Learning Gain (SD)	Average Normalized Gain (SD)	Instructional Time (mins) (SD)
Pilot	Tutored Practice	14.0% (12.8)	34.4% (27.4)	15.2 (6.8)
	Explanation with Practice	5.4% (14.4)	19.7% (32.0)	12.9 (6.0)
	Freeform Explanation with Practice	14.6% (15.2)	35.0% (32.4)	18.5 (7.8)
Study 1	Tutored Practice	12.7% (20.9)	38.3% (42.5)	11.2 (4.3)
	Explanation-only	13.9% (21.3)	40.6% (44.3)	15.0 (5.0)
Study 2	Tutored Practice	17.4% (17.2)	39.5% (30.7)	13.4 (4.3)
	Explanation with Practice	14.1% (15.2)	35.6% (29.7)	18.6 (6.0)
	Analogy with Practice	10.3% (19.3)	29.8% (38.6)	17.0 (7.5)
Study 3	Tutored Practice	11.4% (15.7)	37.2% (35.9)	13.9 (4.6)
	Explanation with Practice	14.2% (18.1)	37.6% (37.9)	17.4 (5.4)
	Example Study with Practice	10.9% (14.5)	39.7% (35.6)	14.4 (5.4)
Study 4	Tutored Practice	11.6% (15.5)	33.3% (31.5)	15.7 (4.8)
	Adaptive Self-Explanation	11.7% (17.2)	32.8% (39.9)	17.9 (4.3)

Chapter 5: Learning Curve Analysis

One of the benefits of using computer-based tutoring systems to conduct this research is the large amount of data generated as students use the tutors. This data enables an in-depth analysis of what happens during learning. All the results presented in Chapter 4 dealt with normal pre to posttest learning measures. While important, those data tell only part of the story. In this chapter, I describe what happens *during* instruction. By looking at the data logs and the learning gains, I can begin to understand not just *what* students know but *how* they learned it. I begin by looking at the overall learning curves for each of the four studies and then a breakdown of learning by knowledge component.

Learning curves provide a glimpse into the rate that students learn. They provide insight into the learning process and provide a more complete picture regarding how students learn a particular unit or individual knowledge components. In order to interpret the learning curves, it is important to begin with a discussion regarding how the curves are generated. To illustrate, consider the data in Table 19, which shows a tutor containing 12 questions and covering 3 knowledge components (A, B, and C). "Opportunity count" is the number of times that a student has completed a step for the associated knowledge component, and "average correct" is the percent of students who answered that question correctly on their first attempt (for the purposes of the learning curves, hint requests are counted as incorrect attempts).

Table 19. Hypothetical data for the first two opportunities for each of the six knowledge components.

Question #	KC	Opportunity Count	Average Correct on First Attempt
Q1	A	1	60%
Q2	B	1	40%
Q3	C	1	80%
Q4	B	2	42%
Q5	B	3	44%
Q6	B	4	50%
Q7	C	2	85%
Q8	A	2	70%
Q9	A	3	71%
Q10	C	3	85%
Q11	A	4	76%
Q12	C	4	85%

To create the learning curve, the average of averages is calculated for each opportunity, collapsing across knowledge components. Thus, for *Opportunity 1*, one would average the values for Q1 (KC A, 60%), Q2 (KC B, 40%), and Q3 (KC C, 80%) and report an average correct for Opportunity 1 of 60%. Similarly, for *Opportunity 2*, one would average the values for Q4 (KCB, 42%), Q7 (KC C, 85%) and Q8 (KC A, 70%) and report an average correct for Opportunity 2 of 65.7%. This process is repeated for all subsequent opportunity counts and when plotted, results in a learning curve (Figure 16).

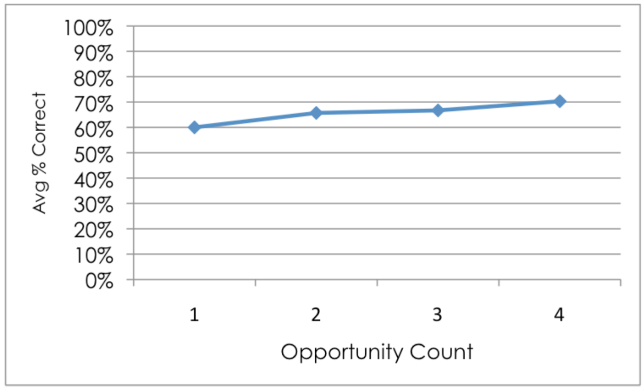
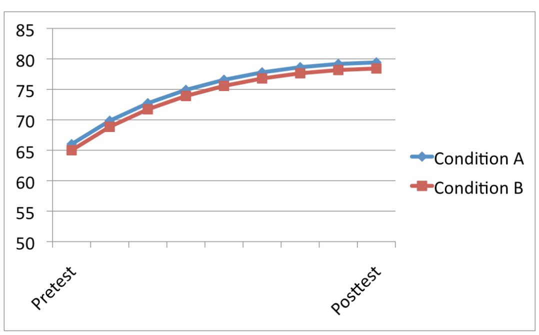
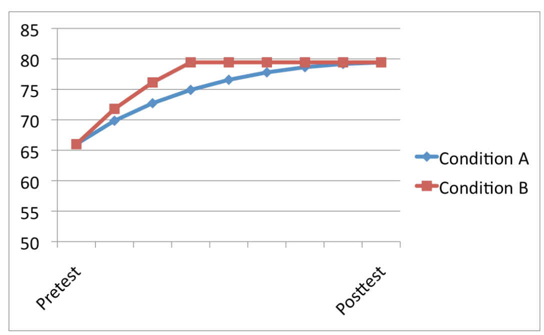


Figure 16. Hypothetical data from Table 19 to illustrate the process of building a learning curve.

Of primary interest is investigating whether students in different conditions follow similar learning curves or not (Figure 17). While the normal learning gains reported in Chapter 4 show that students make the same improvement between the pre and posttests between conditions, those data do not address the details regarding the temporal aspects of learning. For example, do students in different conditions follow the same pattern of learning (Figure 17a) or does one condition reach asymptote early in the process (Figure 17b).



17a



17b

Figure 17. Examples of hypothetical learning curves. In the first (17a), there is no difference between the two conditions. In the second (17b), students in hypothetical condition B learn more from the initial learning opportunities and quickly reach asymptote.

To address these questions, I built learning curves for each of the four studies grouping students by condition (Figure 18-22). The data includes all student responses while using the tutor, but depending on the study, the opportunity counts reflect different activities. For Study 1 (Figure 18), the tutored practice (green line) reflects students' accuracy with selecting the correct article, and the explain-only (pink line) reflects student accuracy with selecting the correct explanation. For Studies 2 and 3, due to the interleaving of item-dependent and matched-practice items, there is variation of task within each condition. Recall that in these studies, half of the tutored items were matched practice items while half were varied depending on condition (e.g., self-explanation, analogy, or worked example). Matched practice items were the article selection items that students in all conditions completed during tutoring. For five of the six knowledge components, the matched practice items were opportunity counts 3, 5, 7, 9 and 10. For the sixth KC ("Use no article if the noun is general and non-count"), the matched practice items were opportunities 2, 4, 6, 9, and 10. The difference in order was due to an interface design decision that limited the number of sentences per screen to five. Thus, in Study 2, the analogy with practice items and self-explanation with practice items reflect student accuracy for selecting the correct analogous sentence or explanation in addition to article selection. In Study 3, since students were only asked to study the worked examples and did not have to make an answer choice, the worked example with practice condition (blue line) represents the percent of students requesting a hint. Finally, in Study 4, since both the adaptive self-explanation and tutored-practice conditions asked students to first select the correct article, both lines represent student accuracy with the article selection task.

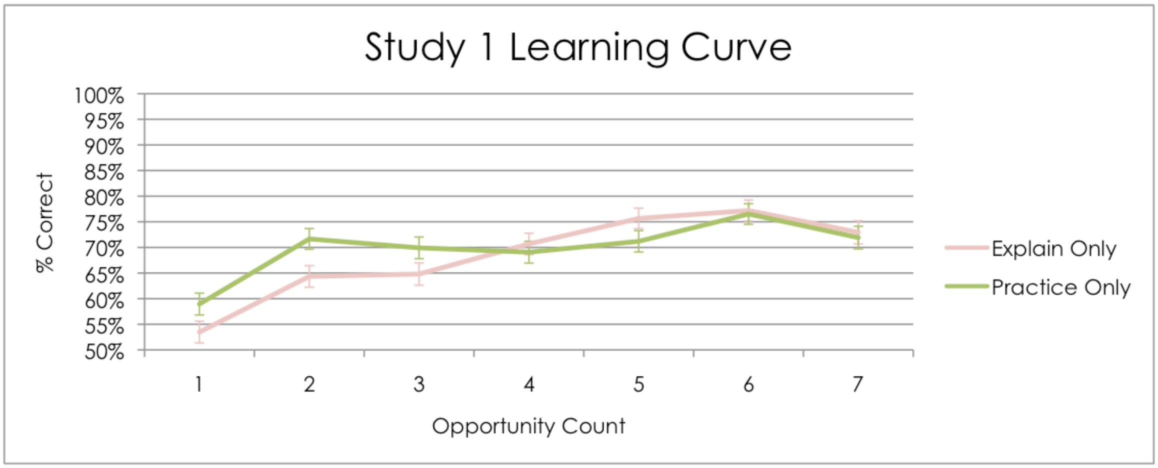


Figure 18. Learning Curve for Study 1

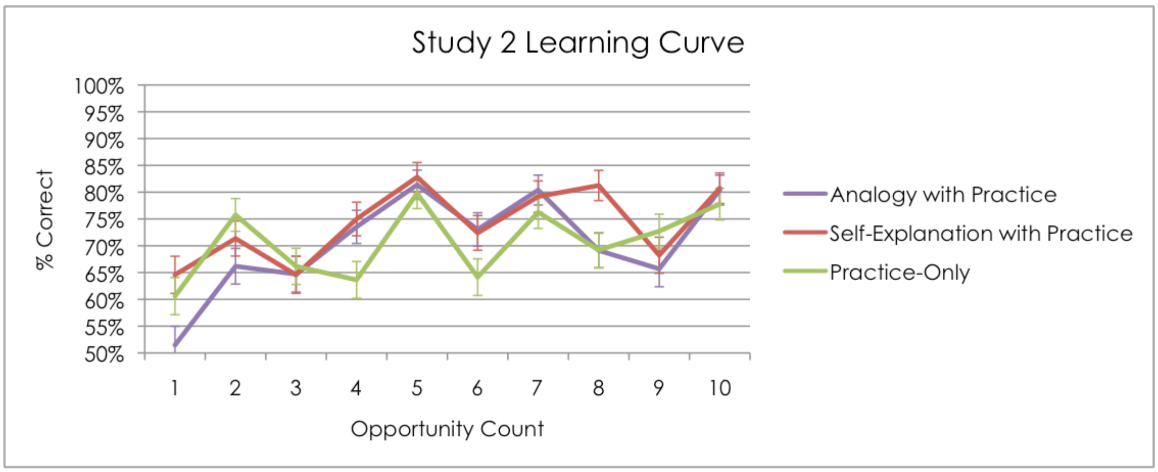


Figure 19. Learning Curve for Study 2.

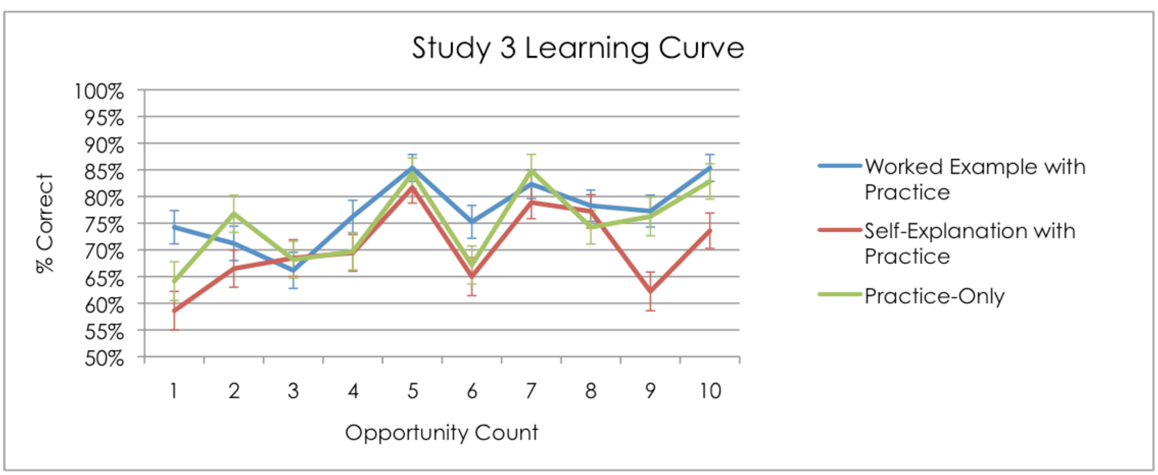


Figure 20. Learning Curve for Study 3.

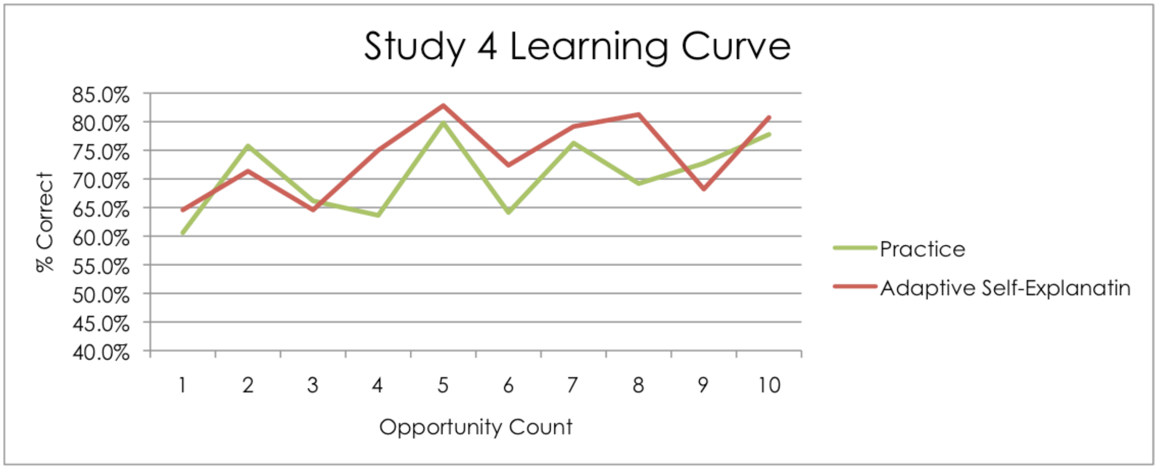


Figure 21. Learning Curve for Study 4.

The main conclusion from the learning curves is that across all conditions, students are making improvement (the general slopes are positive) and overall, there are no consistent differences between conditions (for the most part lines rise and fall at similar intervals). In order to more systematically test whether there is a difference between conditions, I calculated the difference between conditions at each opportunity. If the students were learning at different rates, one would expect the difference between conditions to grow larger in the beginning until the faster learning condition reached asymptote and then the differences to shrink while the slower learning condition caught up. Since the primary conditions of interest were tutored practice (article selection with feedback) and self-explanation with practice, the following analysis only includes data from these two conditions (Study 2 and Study 3 combined). In addition to control for task variability, I only include the matched practice items. A repeated measures ANOVA with percent correct at each of the five opportunities as the dependent measure and condition as the independent measure shows no significant interaction between opportunity count and condition ($F(4, 123)=2.0$, $p = 0.10$) (Table 20).

Table 20. Percent correct for matched practice problems (Study 2 and 3). Students in both the practice and self-explanation conditions show no difference in learning across multiple opportunity counts suggesting that they are learning at the same rate while using the tutors.

Matched Practice	Practice	SE with Practice	Difference
1	67%	67%	1%
2	82%	82%	0%
3	81%	79%	2%
4	74%	65%	9%
5	80%	77%	3%

In addition to overall learning, I was also interested in whether if there was a difference between practice and prompted self-explanation for any of the individual knowledge components. In order to examine interactions between condition and individual knowledge component, I built learning curves for each of the six knowledge components covered in the tutor.

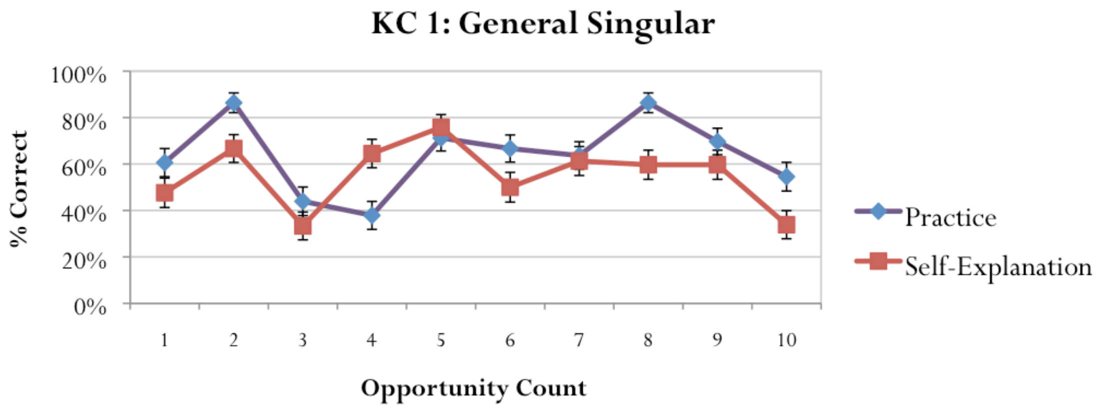


Figure 22. Learning curve for Knowledge Component 1: If a noun is general and singular, use “a” or “an”.

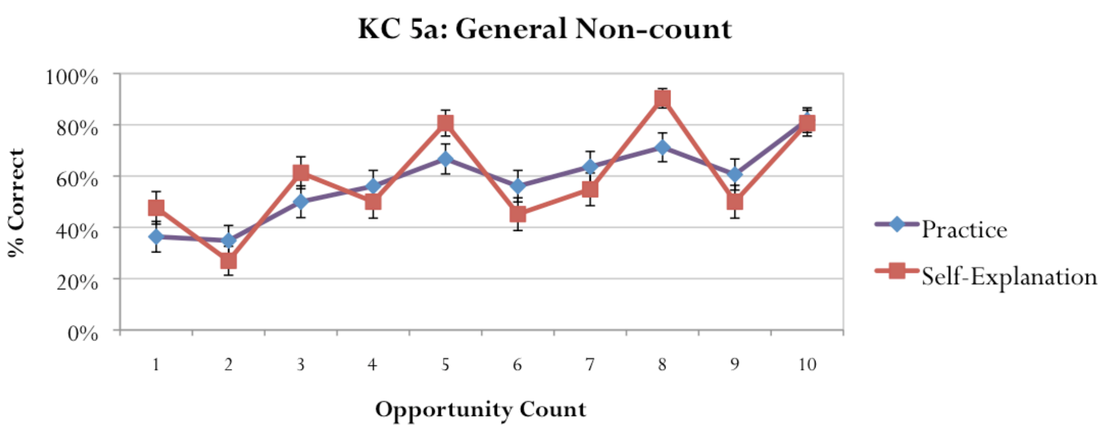


Figure 23. Learning curve for Knowledge Component 5a: If a noun is general and non-count, do not use an article.

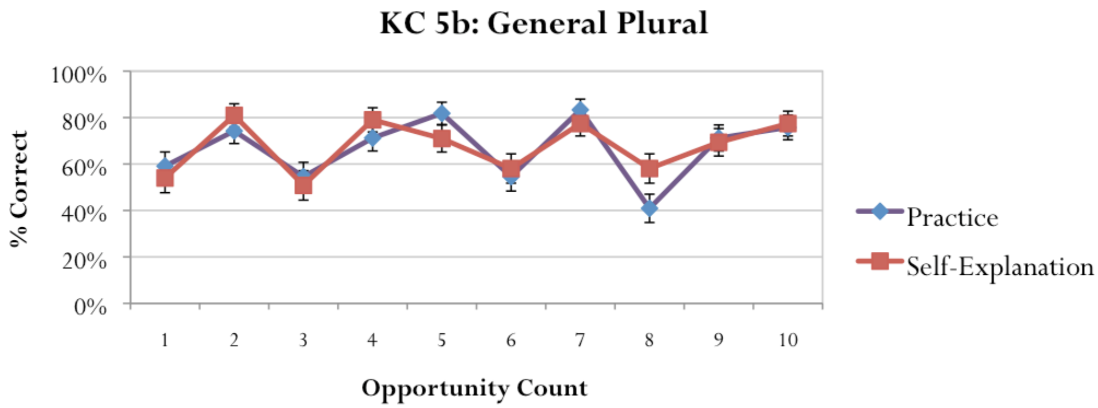


Figure 24. Learning curve for Knowledge Component 5b: If a noun is general and plural, do not use an article.

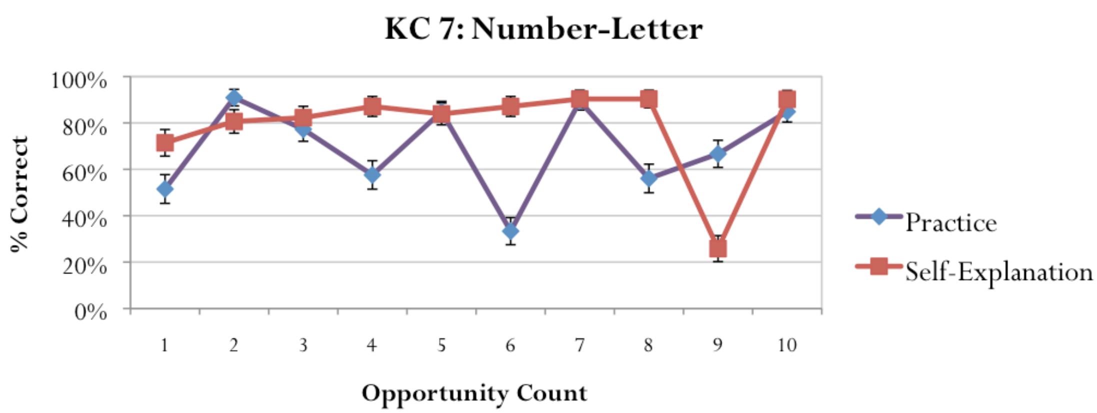


Figure 25. Learning curve for Knowledge Component 7: If a noun is a single letter or number, use "a" or "an".

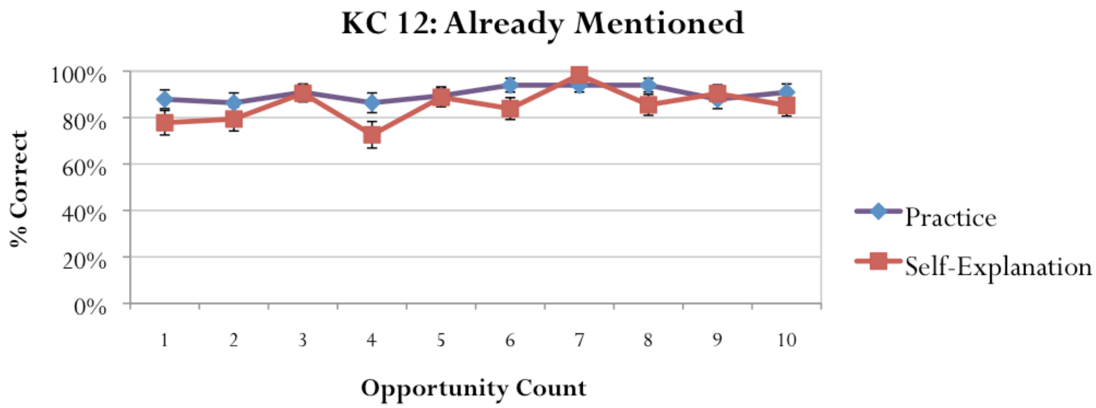


Figure 26. Learning curve for Knowledge Component 12: If a noun has already been mentioned, use "the".

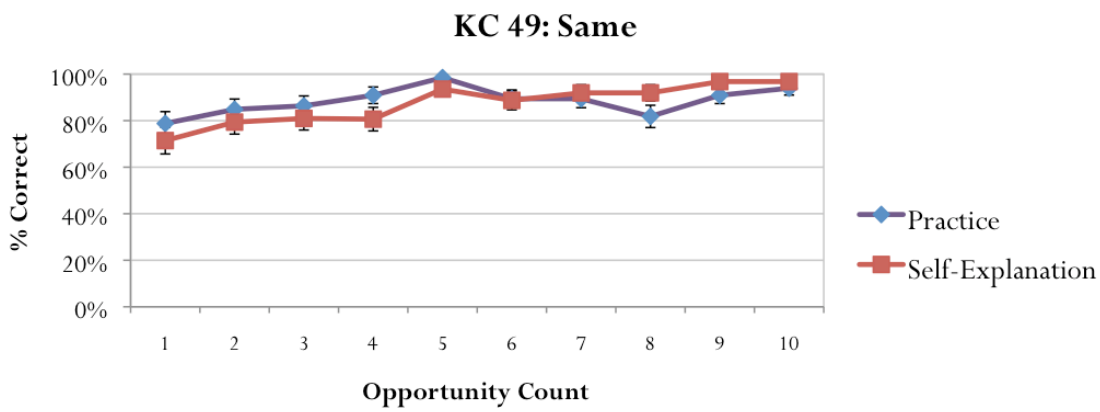


Figure 27. Learning curve for Knowledge Component 49: If a noun is modified with the word "same", use "the".

Unlike the overall learning curves, the individual knowledge component learning curves (Figures 22-27) suggest that, at least for certain knowledge components, there is a difference between practice and prompted self-explanation. The most striking difference between conditions is seen in the number-letter knowledge component (KC 7): If a noun is a single letter or number, use "a" or "an" (Figure 26). For this knowledge component, there are large discrepencies for both opportunity 6 and opportnity 9. The target sentence for opportunity 6 was *The word "hatch" has _an_ "H" at the* *beginning and at the end.* Students in the self-explanation condition were given the correct answer and chose the best explanation while students in the practice condition chose the article that best completed the sentence. At Opportunity 9 (*I'm worried about my son. I think he may get* __an__ *"F" in his history class*), students in both conditions selected the article. I hypothesize that the reason students in the practice condition perform so poorly on Opportunity 6 is that they initially have the shallow rules:

> *If "a" or "an" should be used and the noun starts with a consonant* *letter*, *use "a".*
> *If "a" or "an" should be used and the noun starts with a vowel* *letter*, *use "an"*

rather than the correct rules:

> *If "a" or "an" should be used and the noun starts with a consonant* *sound*, *use "a".*
> *If "a" or "an" should be used and the noun starts with a vowel* *sound*, *use "an".*

Thus, when students are asked to select the correct article, they apply their shallow rule and since "H" is a consonant, they incorrectly select "a". Data support this hypothesis and show that of the 66 students using the practice tutor, 67% were incorrect on their first attempt. Of those, 89% (39 students) chose "a". Students in the self-explanation condition perform very well on this question with 85% correctly selecting "The noun is a single letter or number" as the correct explanation. I propose that because a student needs to apply both rules (*If a noun is a single letter or number, use "a" or "an"* and *If a noun begins with a vowel sound, use "an"*) to correctly select the article, only students in the practice condition are likely to repair or replace their shallow/buggy rule with the correct rule. Conversely, since explaining the sentence only requires students to use the rule *If a noun is a single letter or number use "a" or "an"*, it is unlikely that self-explaining this sentence helps students acquire the important distinction between spelling and phonetics. Furthermore, there is evidence that when students do receive feedback on the article selection task that "a" is incorrect, they do repair their shallow rule. The next opportinuty that presents a phonetic and spelling mismatch for this knowledge componnet is Opportunity 9 (*I'm worried about my son. I think he may get* __an__ *F in his history class.*) Performance data for this opportunity shows that students in the practice condition do significantly better at selecting *an* instead of *a* (McNemar test, $p<0.001$). However, because the features/rules in the self-explanation menu do not specifically address the shallow rule, students in the self-explanation condition do not repair the shallow rule at Opportunity 6 and thus make the same type of error on Opportunity 9 that students in the practice condition made on Opportunity 6 (68% of SE students select "a" as their first choice on Opportunity 9).

In addition, it is important to note that students do not necessarily require feedback to notice and repair their shallow rule. Data from the example study condition (Study 3 only) suggest that feedback is not the essential characteristic. Students in the example study condition were asked only to study Opportunity 6 and given two options: asking for a hint or checking a box to indicate that they were finished studying. The data reveal a large spike in hint requests (42% of students ask for a hint on opportunity six, compared to an average of 24% on previous study problems). While a hint request may signify a request for either instruction or feedback, the large jump suggests uncertainty in the use of "an" in the sentence "The word "hatch" has <u>an</u> "h" at the beginning and at the end." Interestingly when presented with Opportunity 9, students in the example study condition perform more like students in the practice condition than those in the explanation condition; that is, they are more like to get the problem correct. This pattern of behavior suggests that students in the example study condition are also noticing and repairing their shallow rule, even though they don't receive explicit feedback on the earlier trial.

These data provide preliminary evidence that, at least for a very specific and narrow case, the practice and self-explanation tutors are affecting students in different ways. More broadly, these data lead to insights and pointers to future work. For example, they provide evidence that much care needs to be taken when designing menu based self-explanation prompts because by identifying the features to include in the menu, one is implicitly saying that all other features are not important.

Chapter 6: Individual Differences

While the study results consistently showed that there were no differences between instructional conditions on article selection knowledge gain, it is important to ask whether the aggregated group scores reflect how individuals are performing. Thus, in the next two sections, I investigate two potential sources of individual differences, a student's first language and prior knowledge, in order to test whether the results described in Chapter 4 hold for subgroups of the population.

6.1 Effects of First Language

According to second language acquisition research, learning articles should be particularly difficult for students whose native language does not use articles or for those whose article systems are quite different (Luk & Shirai, 2009). Several studies have shown that students who have an article system in their L1 that is similar to that of English perform better on written tests than students whose L1s differ from English (Odlin, 1989). In particular, students appear to transfer language patterns from their first language to English (e.g., Jarvis, 2002; Ringbom, 1987). In order to investigate effects of first language on instruction and potential interactions between first language and instructional condition, I replicated previous analysis adding L1 as a covariate.

Student Population

One of the limitations of *in vivo* experimentation is little control over the specifics of the studied population. The students participating in the studies were the students enrolled in the classes and thus the population was subject to the variability of enrollment. For example, while in total, 22 native languages were represented in the sample, approximately two-thirds of the students spoke Arabic, Chinese, or Korean as their native language (Table 21). Fortunately, however, this sample does facilitate nice theoretical comparisons because each language uses articles in a different manner from each other and also differently from English. In Arabic, the definite article (*al-*) is a prefix and attaches to the word it is modifying. Similarly, the indefinite article is not marked with a distinct word but instead with a suffix (Ryding, 2005). Relative to English, the definite article is used much more frequently in Arabic. For example, in Arabic, the definite article is used to indicate general classes of persons or things (e.g., *Apples are delicious*), abstractions (e.g., *Truth is stranger than fiction*), and for actions or states (e.g., *After graduating, she moved to California.)* (Abboud and McCarus, 1983). Thus, assuming direct translation, one might expect native Arabic speakers to overuse *the*. In contrast, Chinese has a very different article system. There is no definite article but there is a word meaning "one" which can be used to indicate indefiniteness. While definiteness is not frequently marked, it can be indicated through word order or the use of demonstratives (Robertson, 2000). For example, in the example text below, in (a), the word *rén* appears in the topic position indicating that it is definite; whereas in (b), *rén* occurs after the verb indicated that it is indefinite. While different from the English system, Chinese does have other demonstratives that can be used to signify definiteness (*zhèi* for "this" and *nèi-* for "that") and *yi-* or "one" for indefiniteness**.** Previous studies have shown that Chinese L1 students do have a tendency to omit articles, and there is also evidence that they substitute *this*, *that*, and *one* when native speakers would use *the* or *a/an* (Robertson, 2000). Finally, Korean is different from both Arabic and Chinese in that it has neither a definite nor indefinite article thus removing any effects from L1 tranfer (Ionin, 2006). For a summary of the types of article used in Arabic, Chinese, and Korean, see Table 22.

a. rén lái le
person come PFV/CRS[4]
'The person(s) has/have come.'

b. lái-le rén le
come-PFV person CRS
'Some person(s) has/have come.'

Example from Li and Thompson (1981), pg. 20

Table 21. Native Languages of Students. Students came from a variety of L1 backgrounds, with the majority of students speaking Arabic, Chinese, or Korean.

L1	Study 1	Study 2	Study 3	Study 4	Total
Arabic	30	32	36	62	160
Arabic & French	1				1
Bambara			1		1
Chinese	17	18	16	21	72
Chinese, Taiwanese			1	1	2
Farsi				1	1
French	1	1		2	4
Georgian		1	1		2
German	2	1			3
Hebrew		2			2
Italian				1	1
Japanese	6	5	6	6	23
Japanse	1				1
Korean	20	18	9	11	58
Polish	1				1
Portuguese	1	1	4	1	7
Russian	3	1			4
Samoan		1			1
Spanish	4	7	10	3	24
Suundi			1		1
Taiwanese		1		2	3
Thai	5	5	3	1	14
Turkish	5	4	4	1	14
Vietnamese	1			2	3
Unknown	8		1	10	19
Grand Total	106	98	93	125	422

Table 22. Article use by language. The three most common first languages (Arabic, Chinese, and Korean) each have a different article system from English (Dryer, 2008).

Language	Definite Article?	Indefinite Article?
Arabic	Yes	Yes
Chinese	No	Yes
Korean	No	No

In order to test whether the data reflect L1 transfer effects as well as investigate any interactions between L1 and instructional condition, I conducted a repeated measures learning analysis using article pretest and posttest as the dependent measures, condition as an independent variable and L1 as a covariate. Since Study 2 and 3 were identical in form, I combined the data from both studies and only included students whose first languages were Arabic, Chinese, and Korean. Because the goal of this analysis was to investigate effects of first language, I chose to exclude the student listed both Arabic and French as his first languages.

Table 23. Pretest Score by Condition and Native Language. Native Languages representation was uniformly distributed across conditions and there were no significant differences among pretest scores.

Native Language	Practice Condition Average Pretest Score	Explanation with Practice Condition Average Pretest Score	Overall Average Pretest Score
Arabic	65.7% (n=26)	65.3% (n=24)	65.5% (n=50)
Chinese	66.7% (n=7)	66.0% (n=13)	66.3% (n=20)
Korean	57.9% (n=7)	65.6% (n=12)	63.0% (n=19)

Before conducting the analysis, I first checked for equivalence between groups. There was no difference among pretest scores ($F(2,89=0.148$, $p=0.86$) and students from each L1 were uniformly distributed across conditions ($\chi^2=2.31$, df=2, $p=0.32$) (Table 23). While all L1s demonstrated significant pre to posttest improvement, there was no interaction between first language and learning gains (test*L1, $F(2, 83)=3.15$, $p=0.11$) and no three-way interaction between test, L1 and condition (test*L1*condition, $F(2,83)=1.94$, $p=0.15$).

Table 24. Average pretest and posttest score by native language.

Language	Pretest	Posttest
Arabic	65.5% (17.7)	78.2% (13.7)
Chinese	66.3% (23.6)	77.9% (16.3)
Korean	63.0% (21.6)	85.2% (8.9)

6.2 Prior Knowledge

As described in the motivation for Study 4, previous work has found that self-explanation can be more beneficial for students with low prior knowledge than for those with high prior knowledge. In this section, I explore this question with the data collected from across the four studies. In their work using self-explanation to teach reading comprehension strategies, McNamara and colleagues (2004) found that prompting students to self-explain was more beneficial for students with low prior knowledge than those with high prior knowledge. However this finding was only for text-based comprehension questions and did not hold for inference questions that required more domain-specific prior knowledge. They argue that self-explanation improved comprehension by encouraging students to actively process the material, resulting in a better understanding of the material and a reduction in misconceptions. To see if self-explanation is more beneficial for students with low prior knowledge in this domain, I first did a median split on pretest scores and then replicated the repeated measures analysis adding prior knowledge as a covariate. Unlike in McNamara's results, I found no interaction between prior knowledge and condition (condition*prior knowledge, $F(5, 402)=0.48$, $p = 0.79$). However, like McNamara's findings, data show that students with low prior knowledge made significantly larger gains than those with high prior knowledge (prior knowledge, $F(1, 402)=97.74$, $p < 0.001$).

Table 25. Average pretest and posttest scores for low and high prior knowledge students.

	Low Prior Knowledge		High Prior Knowledge	
Condition	Pretest	Posttest	Pretest	Posttest
Study 1				
Practice	56.1%	71.7%	80.4%	82.7%
Explain-Only	59.1%	74.3%	81.3%	80.3%
Study 2				
Practice	44.7%	71.5%	81.0%	85.7%
Self-Explain with Practice	52.2%	76.3%	77.8%	83.3%
Analogy with Practice	55.2%	71.4%	77.5%	80.9%
Study 3				
Practice	55.6%	76.4%	81.7%	86.9%
Self-Explain with Practice	54.4%	73.7%	84.4%	87.5%
Example Study with Practice	61.7%	79.4%	81.9%	87.0%
Study 4				
Practice	54.0%	73.9%	79.9%	84.1%
Adaptive Self-Explain	53.2%	73.8%	79.4%	81.1%

In order to investigate if this effect would persist beyond the immediate posttest, I used the long-term retention data collected as part of Study 4. These results were very promising and showed the gains that the students with low prior knowledge made during tutoring persisted on both the one-week and two-month retention measures (Figure 28). While students with low prior knowledge do not exactly "close the gap", after tutoring, they are performing closer to their high prior knowledge peers and continue to do so even two months after instruction.

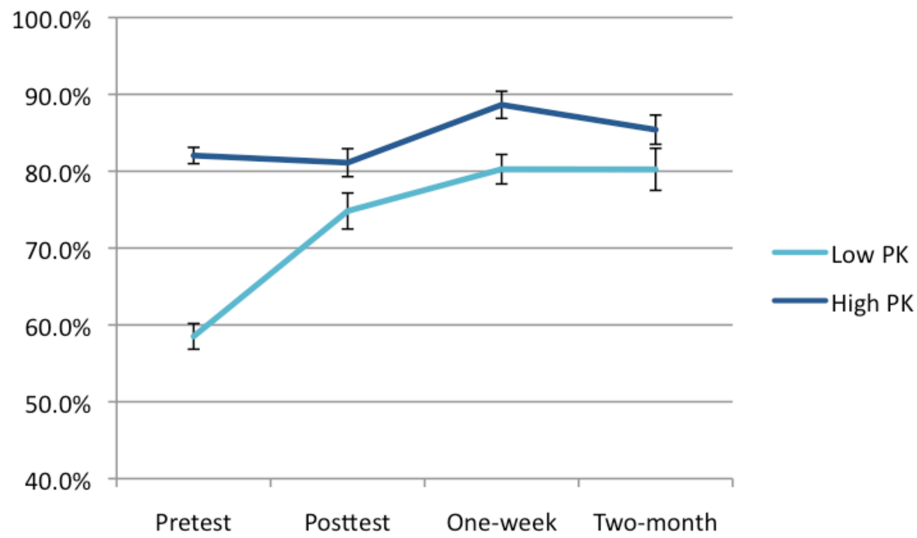


Figure 28. Students with low prior knowledge benefited more from tutoring than students with high prior knowledge.

Overall, all conditions lead to equal learning gains across student first languages and prior knowledge, and there is no evidence to suggest an interaction between tutoring condition and the examined sources of individual differences. However, the data do reveal strong benefits for tutoring in general, especially for students with low prior knowledge. Regardless of condition, students with low prior knowledge made significantly greater gains than those with high prior knowledge, effectively narrowing the achievement gap between the two groups and continued to do well on long-term retention measures.

Chapter 7: Conclusions and Contributions

7.1 Conclusions

This series of studies sought to investigate the claim that self-explanation is a domain general strategy. One main and consistent result is that self-explanation is beneficial in that it leads to increased posttest performance. This finding replicates others in the second language acquisition community that show benefits for metalinguistic instruction. Prior to this work, the results from previous studies in other domains as well as pedagogical recommendations like those in the 2007 IES Practice Guide suggest that self-explanation is a strategy that is generalizable across domains and learner populations. However, the studies presented here show that whether the self-explanation effect is generalizable or not depends on the metric being used. The results from Study 1 and 4 show that self-explanation does generalize in that it leads to an increase in declarative knowledge. Students in the self-explanation condition (Study 1) and adaptive self-explanation condition (Study 4) showed greater improvement on the declarative knowledge (explanation) measure than students in the tutored practice condition. However, if one looks at whether self-explanation helps students achieve their pedagogical goals of successful article use, there is no evidence that the benefits of self-explanation generalize to second language grammar learning. Across all studies, students in the self-explanation conditions demonstrated significant learning gains on the article measure task, but there was no difference between students in the tutored practice condition and those in the self-explanation conditions. Furthermore, across all studies, explanation required extra instruction time, thus it is inefficient compared to the tutored practice condition. Broadly speaking, the results from these studies suggest that future practice guides should strive to make boundary conditions more clear and make explicit mention of the costs (in this case, added instructional time) associated with the various instructional strategies.

Learning curve analyses support the conclusion that students in different conditions are following similar learning curves. There is no evidence that students in the self-explanation condition are reaching asymptote quicker than those in the practice condition. If anything, there seems to be an advantage for the practice condition. In a detailed analysis for Knowledge Component #7 (*If a noun is a single letter or number, use "a" or "an"*), performance data show that students were much better at repairing a preexisting shallow rule when asked to practice (select the article) than when asked to explain the key feature.

These results suggest that the role declarative knowledge plays in the problem-solving process determines the utility of self-explanation. Mainly, tasks for which novices do not require declarative knowledge or for which declarative knowledge is not a pedagogical goal will not produce the self-explanation effect, while those that require declarative knowledge will benefit from self-explanation prompts over practice alone. In the case of English articles, successful article production does not require explicit declarative knowledge of the rules (and may be more fluent without the use of those rules) and being able to explain article rules is not a pedagogical goal. Thus, prompting students to self-explain provides no additional benefit over regular practice exercises with feedback and hints. In their work in geometry, Aleven and Koedinger (2002) found a similar pattern: prompting students to self-explain did not improve scores on items for which procedural knowledge was sufficient to get the correct answer, and the benefit for self-explanation was only seen on transfer items that required a more explicit representation of the knowledge components. Problems that had a numeric solution could be solved either by retrieving and applying the relevant declarative knowledge or via procedural problem solving knowledge. In contrast, transfer problems (like those that asked students to determine if enough information was present to solve the problem) required declarative knowledge to be solved correctly. In their analysis, results showed a benefit for self-explanation for transfer problems that required declarative knowledge. However, there was no benefit for self-explanation over practice for problems for which declarative knowledge was optional. The results from my studies of article learning also follow this pattern. More specifically, prompting students to self-explain article use did not lead to greater learning over a traditional practice environment and correct English article use does not require declarative knowledge. When examined together, these studies suggest that it is knowledge-type, not domain that determines the effectiveness of self-explanation. To fully examine this hypothesis, future work should continue to investigate new domains that differ with respect to the role of declarative knowledge.

Another way to explain why self-explanation is beneficial for some, but not all, tasks is to understand which step in the problem-solving process is supported by self-explanation. To illustrate, consider the following domain-general knowledge-application process: (1) Set goal, (2) Identify relevant problem features, (3) Identify corresponding rules or productions, and finally (4) Execute rule or production. While this 4-step process is domain independent, I hypothesize that the domains differ with respect to the relative difficulty of each step. For example, in geometry, each of these steps present challenges. However, in second language article learning, Step 2, identifying the relevant features of the sentence seems to be the most difficult. Not only must students identify the important features (e.g., the noun has already been mentioned) but they must ignore irrelevant features as well (e.g., *dog* is a singular, count noun). However, once students identify the relevant features, determining and executing the rule (Steps 3 and 4) are relatively simple. These differences suggest another hypothesis for why self-explanation is not as effective in English article learning. Namely, self-explanation targets the mapping of features to rules and productions (Step 3), thus when this is a key source of difficulty, as may be the case in domains like geometry, self-explanation is an effective and efficient strategy. However, when other steps are more problematic, such as identifying the relevant features (Step 2), self-explanation will not be as effective. In such cases, letting more implicit learning mechanisms discern on feature patterns may be just as effective and more efficient.

7.2 Limitations

Decisions made throughout the course of this work resulted in limitations. Those that have the biggest potential impact on the generalizability of the results include student population, knowledge component selection, and manner of instruction (e.g., limited instruction time and sentence-level format).

This work was conducted through collaboration with the Pittsburgh Science of Learning Center's English LearnLab. While this collaboration was extremely beneficial and facilitated feedback from instructors and *in vivo* experimentation, the students enrolled in the course may not be typical learners. Specifically, students were all adult English language learners who were relatively advanced in their English language development and had all been introduced to English articles previously in their course of study. While important to note that this may reduce generalizability to other populations, the number of adult English language learners is very large. According to the American Association of Intensive English Programs, in 2008, approximately 60,000 students were enrolled in an Intensive English Program in the United States alone.

Another potential limitation is that due to limited class time, only a subset of the knowledge components in the model were covered in the tutor. While much care was taken to choose knowledge components that covered a range of frequency within written corpora, it is possible that self-explanation could be beneficial for certain knowledge components that were not chosen to be part of the tutoring system. Furthermore, self-explanation may become more beneficial as the number of knowledge components increase and with them, the cognitive demands of differentiating between relevant features and rules.

Finally, in order to insure proper alignment between instruction and assessments, all tutor and test items were presented as single cloze statements. Perhaps modifying the form to include complete paragraphs or essays would alter the results. For example, self-explanation may be more useful when the body of writing is more complex and thus the features for making article decisions are less salient. Finally, moving to a larger body of text would increase the ecological validity as students are often asked to produce more than sentence-length utterances.

7.3 Contributions

This thesis combines human-computer interaction and learning sciences methods to make contributions to the learning sciences, second language acquisition, and tutoring system communities.

Contributions to the Learning Sciences

With respect to the learning sciences, this work cautions against broad generalizations for instructional strategies and suggests that strategies should be aligned to specific target knowledge and not domain. For example, instead of labeling instructional strategies for a specific domain like physics learning or grammar learning, instructional strategies should be chosen with respect to knowledge goals we want students to achieve (e.g., declarative knowledge, procedural knowledge, etc.). Importantly, this work encourages the field to shift the way instructional effectiveness is described. In lieu of claiming that instruction is better suited for one *domain* versus another, these results suggest that the discussion should shift to the level of knowledge type. This perspective is shared and serves as the basis for the KLI framework (Koedinger, et al., submitted). More specifically, this work contributes to the learning sciences by showing that there are limits to the benefits of self-explanation. While prompting students to self-explain does increase posttest performance, results from all four studies show that self-explanation requires significantly more time to complete than tutored practice. Additionally, while self-explanation leads to increased declarative knowledge, for the English article domain, this does not transfer to the more pedagogically relevant skill, the article selection task. Finally, in addition to prompted self-explanation, individual studies also showed no advantage for analogical comparisons or worked example study over tutored practice.

Contributions to Second Language Acquisition

The main contribution of this work to second language acquisition theory is further data supporting the benefits of explicit instruction in grammar learning and demonstrating that a broad spectrum of explicit learning conditions are beneficial to students (Table 26). This dissertation contributes to second language acquisition through systematically investigating a number of explicit strategies and comparing their relative effectiveness and efficiency for learning the English article system. While all forms of instruction were equally effective, there were differences with respect to efficiency. The issue of instructional efficiency in second language acquisition is understudied and this work highlights the importance of considering both instructional effectiveness and the amount of time required to achieve it.

Table 26. Instructional conditions covered a wide spectrum of explicit forms of instruction.

	Condition	Form of Explicit Instruction
Less Explicit	Worked Example Study	Metalinguistic explanations in hints
	Practice	Immediate corrective feedback on article selection and metalinguistic explanations in hints
	Analogical Comparisons	Immediate corrective feedback on analogy selection and metalinguistic explanations in hints
More Explicit	Prompted Self-Explanation	Immediate feedback on metalinguistic explanations and metalinguistic explanations in hints

In addition, this work demonstrates the benefits of computer-based tutoring systems for second language grammar learning and especially for students with low prior knowledge.

Contributions to Tutoring Systems

Finally, this work also makes contributions to tutoring systems research. I've demonstrated a process for data-driven and experiment-driven tutor design that has produced significant learning gains and consistent adoption in real classrooms. After each stage, I used the data to refine and improve the tutoring systems. An early example was the move from paragraph-level to sentence-level instruction. This modification enabled better alignment and ensured that students received a sufficient number of practice opportunities for each knowledge component.

I have also demonstrated an iterative process for developing and refining a knowledge component model. The model was initially seeded with theoretically driven rules based on the desired student behavior or answer and refined with a more cognitive approach and validated using data collected through the tutoring systems. Using this model for instruction and assessment alignment produced significant learning gains and much improvement over the initial answer-based model.

7.4 Future Work

In addition to empirically testing and addressing the previously mentioned limitations, this dissertation lays the groundwork for several areas of future research. While I argue that the debate needs to shift away from identifying strategies for specific domains and instead align strategies to knowledge goals, to fully test this theory, one would need to compare the same knowledge type across different domains. For example, do more complex knowledge components within second language learning benefit from more complex learning strategies, like self-explanation? Conversely, do simpler knowledge components in math or science not show an effect for complex learning strategies?

This research also suggests that future work should investigate alternative methods for implementing self-explanation into computer-based tutoring systems. As evidenced by the detailed analysis of the letter-number knowledge component learning curve, which showed that the practice condition (but not the self-explanation condition) encouraged students to repair their shallow rules regarding when to use "a" versus "an", deciding what rules or features to include in the self-explanation menu is critical. Additionally, it appears that for some sentences, more than one feature or knowledge component should be identified or explained. Follow-up work could investigate whether benefits of self-explanation appear if students select all features of a noun rather than the most important for making the article decision. For example, using checkboxes, students could identify whether a noun was singular, plural, or non-count, whether the noun was generic or definite, and if definite for which reason(s). This process would provide a more complete representation of a student's knowledge and allow for more specific feedback and scaffolding to be provided. By asking students to choose multiple features, this type of interface would also prevent students from focusing too narrowly on specific features and ignoring others.

In addition, this work raises the question about whether generating the explanation or selecting it from a menu is necessary; what if, instead of selecting the most important features, students were simply given the explanation? As mentioned in the introduction, Gerjets' (2006) study failed to find an advantage for prompted self-explanation. One key difference between that study and other self-explanation studies is the comparison condition. Gerjets compared prompted self-explanation to fully-explained worked examples and thus tested whether *generating* the explanation was beneficial compared to being *given* the explanation. Similarly, work by Presson, et al. (submitted), demonstrated that providing students with explicit rules in response to errors was more beneficial than providing the answer alone. These results suggest that explicit or metalinguistic feedback is beneficial for learning but generating it may take too much time with no associated benefits.

More broadly, these studies motivate future work to test the different theories regarding the boundary conditions of prompted self-explanation. Specifically, the data suggest two theories: (1) The role declarative knowledge plays in the problem-solving process determines the utility of self-explanation, and (2) Tasks for which feature to rule mapping is challenging will benefit from prompted self-explanation. In order to tease apart these two theories, future work should look at the effects of self-explanation in domains that vary in terms of declarative knowledge use and mapping difficulty. Furthermore, in order to differentiate these attributes from those of overall domain categorization, tasks should be chosen from both STEM and non-STEM domains. For example, within second language acquisition, one could compare English article learning (a task that has not shown to benefit from additional declarative knowledge) to a task like regular verb conjugation in Spanish. Conjugating regular verbs in Spanish follows a very specific process that is explicitly taught to students, and presumably, greater declarative knowledge of the process results in better performance. If verb conjugation skills were improved as a result of self-explanation, this would be evidence towards the first theory that the greater role declarative knowledge plays in the problem-solving process, the greater the benefit for self-explanation. Similarly, one could also test the second theory and investigate domains that vary among the dimension of feature-to-rule mapping difficulty. For example, one could administer an assessment that isolates each of the problem-solving steps in order to determine relative difficulty of each step and then compare those results to the observed benefits of self-explanation across tasks. The second theory would predict a correlation between tasks that show benefits for self-explanation and tasks for which the rule mapping step is the most challenging.

References

Abboud, P., & McCarus, E. (1983). Elementary modern standard Arabic. Cambridge University Press.

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.

Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). Example-tracing tutors: A new paradigm for intelligent tutoring systems. International Journal of Artificial Intelligence in Education.

Anderson, J.R. & Lebiere, C. (1998). *The Atomic Components of Thought.* Mahwah, NJ: Erlbaum.

Atkinson, R.K., Derry, S.J., Renkel, A., & Wortham, D.W. (2000). Learning from examples: Instructional principles of the worked examples research. Review of Educational Research, 70, 181-214.

Atkinson, R., Renkl, A., & Merrill, M. (2003) Transitioning from studying examples to solving problems: Combining fading with prompting fosters learning. Journal of Educational Psychology.

Bickerton, D. (1981). *Roots of Language.* Ann Arbor: Karoma Publishers.

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, 221-252.

Box, G. (1979) Robustness in the Strategy of Scientific Model Building. Robustness in Statistics: Proceedings of a Workshop. New York: Academic Press.

Calin-Jagerman, R.J., & Ratner, H.H. (2005). The role of encoding in the self-explanation effect. *Cognition and Instruction*, *23*, 523-543.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, *15*, 357–366.

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

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, Hillsdale, NJ: Lawrence Erlbaum Associates. 161-238.

Clark, R. C., & Mayer, R. E. (2004). E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning. San Francisco: Jossey-Bass.

Cole, T. (2000). The Article Book: Practice Towards Mastering a, an, and the. University of Michigan Press.

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.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning & Verbal Behavior. 11(6), 671-84.

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

Doughty, C. (2001). Cognitive underpinnings of focus on form. In P. Robinson (Ed.), Cognition and Second Language Instruction, 206-257.

Dryer, M. (2008) Definite Articles. In: Haspelmath, Martin & Dryer, Matthew S. & Gil, David & Comrie, Bernard (eds.) The World Atlas of Language Structures Online. Munich: Max Planck Digital Library, chapter 37. Available online at http://wals.info/feature/37. Accessed on October 28, 2009.

Ellis, N. (1994). Implicit and Explicit Language Learning – An Overview. In N. Ellis (Ed.) Implicit and explicit learning of languages (pp. 1-32). London: Academic Press.

Ellis, N. and Sagara, N. (2010). The Bounds of Adult Language Acquisition. Studies in Second Language Acquisition. 32, 553-580.

Ellis, R. (1994). The Study of Second Language Acquisition. Oxford University Press.

Ellis, R., Loewen, S., and Erlam, R. (2006). Implicit and Explicit Corrective Feedback and the Acquisition of L2 Grammar. Studies of Second Language Acquisition. 28, 339-368.

Fuchs, M., Bonner, M., & Westheimer M. (2006). Focus on Grammar 3: An Integrated Skills Approach, Third Edition. White Plains, NY: Pearson Education, Inc.

Gick, M.L., & Holyoak, K.J. (1983). Schema induction and analogical transfer. Cognitive Psychology, 15, 1-38.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. Cognitive Science. 7 (2), pp155-170.

Gentner, D., Loewenstein, J., Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. Journal of Educational Psychology, 95(2), 393-408.

Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. D. (2009). Reviving inert knowledge: Analogical abstraction supports relational retrieval of past events. Cognitive Science, 33, 1343-1382.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction*, *16*, 104-121.

Hake, R. (1998). Interactive-engagement versus traditional methods: a six-thousand- student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66, 64 – 74.

Ionin, T. (2006) This is definitely specific: Specificity and Definiteness in Article Systems. Natural Language Semantics, 14, 175-234.

Jarvis, S. (2002). Topic continuity in L2 English article use. *Studies in Second Language Acquisition, 24*, 387–418.

Kim, H., & Mathes, G. (2001). Explicit vs. implicit corrective feedback. *The Korea TESOL Journal*, *4*, 1–15+

Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), Handbook of Metacognition in Education (pp. 897–964). The Educational Psychology Series. New York: Routledge.

Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J. (2010) A Data Repository for the EDM community: The PSLC DataShop. In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) Handbook of Educational Data Mining. Boca Raton, FL: CRC Press.

Koedinger, K., R., Corbett, A.T., and Perfetti, C. (submitted). The Knowledge-Learning Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Learning. Manuscript submitted for publication.

Lenneberg, E. H. (1967). Biological Foundations of Language. Wiley.

Li, C., & Thompson, S. (1981). Mandarin Chinese: a functional reference grammar. Berkeley: University of California Press.

Long, M., Inagaki, S., and Ortega, L. (1998). The Role of Implicit Negative Feedback in SLA: Models and Recasts in Japanese and Spanish. The Modern Language Journal. 82(3), 357-371.

Lovett, M., Meyer, O., & Thille, C. (2008). Open Learning Initiative: Testing the accelerated learning hypothesis in Statistics. *Journal of Interactive Media in Education*, 2008(1).

Lu, C.F.-C. (2001). The acquisition of English articles by Chinese learners. Second Language Studies, 20, 43–78.

Luk, Z.P.S., & Shirai, Y. Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the Acquisition of Plural –s, Articles, and Possessive ‘s. Language Learning, 59(4), 721-754.

Lyster, R. (2004). Differential effects of prompts and recasts in form-focused instruction. *Studies in Second Language Acquisition*, *26*, 399–432.

Master, P. (1997) The English Article System: Acquisition, Function, and Pedagogy. System, 25(2), 215-232.

Matsuda, N., Cohen, W., Sewall, J., Lacerda, G., & Koedinger, K. R. (2008). Why tutored problem solving may be better than example study: Theoretical implications from a simulated-student study. In B. Woolf et al. (Eds.), *Proceedings of the 9th International Conference of Intelligent Tutoring Systems.* 111-121.

Matthews, P. & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts and procedures as pedagogical tools. Journal of Experimental Child Psychology, 104, 1-21.

McNamara, D.S. (2004). SERT: Self-explanation reading training. Discourse Processes, 38, 1-30.

McNamara, D.S., & Scott, J.L. (1999). Training reading strategies. In M. Hahn & S.C. Stoness (eds.), Proceedings of the Twenty First Annual Conference of the Cognitive Science Society. pp. 387--392. Hillsdale, NJ: Erlbaum.

Mitchell, R., & Myles, F. (2004) Second language learning theories (2nd ed). London: Edward Arnold.

Nagata, N. (1993). Intelligent computer feedback for Second Language Instruction. *Modern Language Journal*, 77, 330-339.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. Language Learning, 50, 417-528.

Odlin, T. (1989). Language transfer: cross-linguistic influence in language learning. Cambridge University Press.

Paas, F., & Van Merrienboer, J. (1994). Variability of worked examples and transfer of geometry problem solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86, 122-133.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., and Metcalfe, J. (2007). Organizing Instruction and Study to Improve Student Learning (NCER 2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Pirolli, P. and Recker, M. (1994). Learning Strategies and Transfer in the Domain of Programming. Cognition and Instruction, 12, 235-275.

Presson, N. and Macwhinney, B. (submitted). Learning grammatical gender: The effects of rule instruction.

Rau, M., Aleven, V., & Rummel, N. (2009). Intelligent tutoring systems with multiple representations and self-explanation prompts support learning of fractions. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A. Graesser (Eds.), Proceedings of the 14th International Conference on Artificial Intelligence in Education, AIED 2009, 441- 448.

Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N.T., Koedinger, K. R., Junker, B., Ritter, S., Knight, A., Aniszczyk, C., Choksey, S., Livak, T., Mercado, E., Turner, T.E., Upalekar. R, Walonoski, J.A., Macasek. M.A., Rasmussen, K.P.:(2005). The Assistment Project: Blending Assessment and Assisting. In C.K. Looi, G. McCalla, B. Bredeweg, & J. Breuker (Eds.) Proceedings of the 12th Artificial Intelligence In Education. 555--562.

Renkl, A., Atkinson, R.K., Maier, U.H., & Stanley, R. (2002). From example study to problem solving: Smooth transitions help learning. *Journal of Experimental Education, 70*, 293-315.

Ringbom, H. (1987). The role of the first language in foreign language learning. Clevedon, UK: Multilingual Matters.

Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. Second Language Research, 16, 135-172.

Rosenbaum, P. & Rubin, D. (1984) Reducing the bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association, 79, 516-524.

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.

Ryding, K. (2005). A reference grammar of modern standard Arabic.

Sao Pedro, M., Gobert, J., & Raziuddin, J. (2010). Comparing Pedagogical Approaches for the Acquisition and Long-Term Robustness of the Control of Variables Strategy. In: Gomez, K., Lyons, L., Radinsky, J. (eds.) Proceedings of the 9th International Conference of the Learning Sciences, pp.1024-1031.

Schumann, J. (1979). The acquisition of English negation by Speakers of Spanish: A review of the literature. In R. Andersen (Ed.), The acquisition and use of Spanish and English as first and second languages (pp. 3-32). Washington DC: TESOL.

Siegler, R. (2002). Microgenetic studies of self-explanation. In N. Garrott & J. Parziale (Eds.), *Microdevelopment: A process oriented perspective for studying development and learning* (pp. 31-58). Cambridge, MA: Cambridge University Press.

Sweller, J. (1988) Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12, 257-285.

Sweller, J., & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2(1), 59-80.

Tarone, E. (1985). Variability in interlanguage use: a study of style-shifting in morphology and syntax. *Language Learning, 35*, 373-404.

Tarone, E. & Parrish, B. (1988). Task-related variation in interlanguage: the case of articles. *Language Learning, 38,* 21-43.

Trafton, J. G., & Reiser, B. J. (1993). The contribution of studying examples and solving problems to skill acquisition. Proceedings of the 15th Annual Conference of the Cognitive Science Society, 1017–1022.

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

VanLehn, K., Jones, R., & Chi, M. (1992). A model of the self-explanation effect. *Journal of the Learning Sciences*, *2*, 1-59.

Williams, J. & Evans, J. (1998). What kind of focus and on which forms? In C. Doughty & J. Williams (Eds.), Focus on Form in Classroom Second Language Acquisition, 139-155.

Wulf, G. & Shea, C.H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. Psychonomic Bulletin & Review, 9, 185-211.

Wylie, R., Koedinger, K., & Mitamura, M. (2009). Is Self-Explanation Always Better? The Effects of Adding Self-Explanation Prompts to an English Grammar Tutor. In N.A. Taatgen & H. van Rijn (Eds.), Proceedings of the 31st Annual Conference of the Cognitive Science Society, 1300-1305. Austin, TX: Cognitive Science Society.

Appendix

Set of Rules and Exceptions from *The Article Book*

Rule	Rule or Exception	Answer
1. Use a or an when a singular count noun is indefinite. (Use ‘an’ before vowel sounds.)	rule	a or an
1b. Never use ‘a’ or ‘an’ with plural nouns.	rule	null
2. Use ‘a’ when a singular count noun is indefinite and followed by a consonant sounds.	rule	a
3. Use ‘an’ when a singular count noun is indefinite and the article is followed by a vowel sound.	rule	an
4. Do not use a or an with non-count nouns.	rule	null
5a. Do not use an article with a non-count noun that is indefinite.	rule	null
5b. Do not use an article with a plural noun that is indefinite.	rule	null
6. Do not use ‘the’ with ‘there + be’ All nouns plural and singular are indefinite if they occur after ‘there + be’	rule	a, an, null
7. Use ‘a’ or ‘an’ for single letters and numbers.	rule	a
8. Use ‘a’ to mean ‘for each’ or ‘per’ when the noun begins with a consonant.	rule	a
9. Use ‘an’ to mean ‘for each’ or ‘per’ when the noun begins with a vowel sound.	rule	an
10. Use ‘the’ with ‘in the morning’, ‘in the afternoon’, and ‘in the evening.’	rule	the
11. Do not use an article for ‘at night’	rule	null
12. Use ‘the’ when the noun has already been mentioned.	rule	the
13. Use ‘the’ when the noun is already known.	rule	the
14. Use ‘the’ when the noun is made definite by a prepositional phrase.	rule	the
14-Exception: Do not use ‘the’ when the prepositional phrase does not make the noun definite.	exception	a, an, null
15. Use ‘the’ when the noun is made definite by an adjective clause or an adjective phrase.	rule	the
15-Exception: Don’t use ‘the’ when the adjective clause or adjective phrase does not make the noun definite.	exception	a, an, null

16. Do not use an article with the names of streets, avenues, roads, lanes or boulevards.	rule	null
17. Do not use an article when generalizing about abstract nouns.	rule	null
18. Do not use an article when generalizing in the plural.	rule	null
19. Do not use an article with the names of universities or colleges.	rule	null
19-Exception: Use 'the' with names of colleges that contain the word 'of'	exception	the
20. Do not use an article with the names of countries, cities or states.	rule	null
20-Exception: Use 'the' in the names of countries that contain the words 'united', 'union', 'kingdom' or 'republic'	exception	the
21. Use 'the' with the superlative degree.	rule	the
22. Do not use 'the' with the comparative degree.	rule	a, an, null
22-Exception: Use 'the' with the comparative degree for double comparatives or when the adjective is a comparison is used as a noun.	exception	the
23. Use 'the' with ordinal numbers and ranking words like 'next' and 'last'.	rule	the
23b-Exception: Do not use 'the' with 'next' or 'last' when referring to specific items like 'next month', 'last Christmas', 'next Tuesday', or 'last year'	exception	null
23-Exception: Do not use an article with ordinal numbers or ranking words when listing ideas:	exception	null
24. Use 'the' when generalizing about an entire class of musical instruments.	rule	the
24-Exception: Use 'a', 'an' or 'the' if the sentence can mean either the general class of instrument or any one of the instruments.	exception	a, an, the, null
25. Use 'the ' when generalizing about an entire class of animals.	rule	the
25-Excpetion: Use 'a', 'an', or 'the' if the sentence can mean either the general class of animals or any one of the animals.	exception	a, an, the, null
26. Use 'the' when generalizing about an invention.	rule	the
26-Exception: Use 'a', 'an' or 'the' if the sentence can mean either the general class of the invention or any one of the inventions.	exception	a, an, the, null
27. Use 'the' with the names of rivers, oceans, seas, and deserts.	rule	the
28. Use 'the' with plural names.	rule	the
29. Use 'the' with family names followed by a noun.	rule	the
30. Do not use an article for the names of single lakes, mountains, or canyons.	rule	null

31. Use 'the' with the names of hotels, motels, theatres, bridges and buildings	rule	the
31-Exception: Do not use an article with the names of halls or hospitals.	exception	null
32. Use 'the' with the names of zoos, gardens, museums, institutes, and companies.	rule	the
33. Use 'the' when the noun is the only one that exists.	rule	the
34. Use 'the' to express the plural of nationalities that have no other plural form.	rule	the
35. Use 'the' with adjectives that act as plural nouns to describe a group of people (Abstract Adjectives)	rule	the
36. Do not use an article for the names of stadiums, malls, or parks.	rule	null
37. Do not use an article with the names of languages or religions with have not been made definite.	rule	null
37-Exception: Use 'the' when the word 'language' is used after the name of the language or when the word 'religion' is used after the adjective for (or name of) the religion.	exception	the
38. Do not use an article with the words 'few' or 'little' if the meaning is especially negative.	rule	null
39. Use 'a' when the words 'few' or 'little' express a positive meaning.	rule	a
39-Exception: Use 'a' when the words 'few' or 'little are used with the words 'only' or 'just'	exception	a
40. Use 'the' for compass directions if they follow prepositions like 'to', 'in', 'on', 'at', or 'from'	rule	the
41. Do not use an article if a compass direction immediately follows an action verb like 'go', 'travel', 'turn', sail', 'fly', walk', or 'move'.	rule	null
42. Use 'the' with large periods of historic time like the 1990's, the Jet Age, the Dark Ages, the Cambrian Period, etc.'	rule	the
43. Do not use an article with 'go to bed', 'go to school', 'go to college', 'go to class', 'go to church', or 'go to jail'. These are multi-word verbs that have become shortened with repeated use.	rule	null
44. Use 'the' in special names and titles (Epithets)	rule	the
45. Use 'the' for body parts that have been struck or touched by an outside object.	rule	the
46. Do not use an article with the names of diseases.	rule	null
46-Exception: Use 'the' with 'the flu, 'the measles', and 'the mumps'.	exception	the

47. Do not use an article when referring to numbers or letters in a list.	rule	null
48. Use ‘the’ with nouns for military institutions: the army, the navy, the air force, the marines, the police, etc.	rule	the
49. Use ‘the’ with the word ‘same’	rule	the

Sample Student Report

Page 1 of 2

Student: Stu_104c73ecdd6094fbe5431c74f8264aff

Sentence	Correct Answer	
A1. I'm glad that you had such _a_ wonderful expereinces here.	<no article>	X
A2. I'll have _the_ same thing that she is having.	the	
A3. I made some coffee. _The_ coffee was too strong.	The	
A4. _<no article>_ oxygen has eight protons.	<no article>	
A5. There was _a_ fire downtown last night.	a	
A6. For vacation, I always go to _the_ same place - Disneyland!	the	
A7. That salesman is not _an_ honest man.	an	
A8. My father doesn't drink _a_ coffee, but he does drink tea.	<no article>	X
A9. This is the first time that I've received _a_ "99" on a test.	a	
A10. Cristina asked to borrow a pen, and I told her _the_ pens were in the top drawer.	the	
A11. _<no article>_ most Americans do no like to eat snails.	<no article>	
A12. Indonesia beings with _an_ "I".	an	
B1. My sister and I have _the_ same color eyes.	the	
B2. I want to buy a house, so I'm trying to save _<no article>_ money.	<no article>	
B3. Are there_<no article>_apples in the refrigerator?	<no article>	
B4. Some planes appeared, and then _<no article>_ planes landed in a field.	the	X
B5. I need to do well on the final exam to get _<no article>_ "B" in this class.	a	X
B6. There are _<no article>_ cars in the parking lot.	<no article>	
B7. I drank _a_ milk yesterday.	<no article>	X
B8. I drove here early this morning, so that I could go back _the_ same day.	the	
B9. I planted a garden, and _the_ garden grew.	the	
B10. France is _an_ European country.	a	X
B11. My name begins with _<no article>_ "L".	an	X
B12. Do you have _an_ umbrella?	an	

Feature
A1. "Wonderful experiences" in a generic, plural noun
A2. The noun "thing" is modified with the word "same".
A3. "Coffee" has already been mentioned.
A4. "Oxygen" is a generic, non-count noun.
A5. "Fire" is a generic, singular noun that starts with a consonant sound.
A6. "Place" is modified with the word "same".
A7. "Honest man" is a generic, singular noun that starts with a vowel sound.
A8. "Coffee" is a generic, non-count noun.
A9. "99" is a number.
A10. "Pens" has already been mentioned.
A11. "Most Americans" is a generic, plural noun.
A12. "I" is a single letter.

B1. "Color" is modified with the word "same".
B2. "Money" is a generic, non-count noun.
B3. "Apples" is a generic, plural noun.
B4. "Planes" has already been mentioned.
B5. "B" is a single letter.
B6. "Cars" is a generic, plural noun.
B7. "Milk" is a generic, non-count noun.
B8. "Day" is modified by the word "same".
B9. "Garden" has already been mentioned.
B10. "European country" is a generic, singular noun that begins with a consonant sound.
B11. "L" is a single letter.
B12. "Umbrella" is a generic, singular noun that begins with a vowel sound.

Rules
Use **the** if the noun has already been mentioned.
Use **the** if the noun is modified with the word "same"
Use **no article** if the noun is general and non-count.
Use **no article** if the noun is general and plural.
Use **a** or **an** if the noun is a single letter or number.
Use **a** or **an** if the noun is general and singular.

IRT Model Results

Using an IRT model, I estimated item difficulty at both pretest and posttest. I hypothesized that items of KCs that were seen frequently in tutoring would have lower difficulty estimates at posttest than pretest.

Step	KC num	KC Frequency	Pretest Estimate	Posttest Estimate	Learning Expected?
form1_1	KC 7	high	0.077	-2.274	Yes
form1_10	KC 1	low	-24.102	-3.552	No
form1_11	KC 1	low	-24.102	-2.274	No
form1_12	KC 1	low	-1.991	-3.552	Yes
form1_13	KC 1	low	-24.102	-24.556	Yes
form1_14	KC 1	low	-24.102	-3.552	No
form1_15	KC 7	high	-24.102	-3.552	No
form1_16	KC 1	low	-24.102	-3.552	No
form1_2	KC 7	high	-1.991	-1.183	No
form1_3	KC 7	high	-1.991	-3.552	Yes
form1_4	KC 2	low	-24.102	-3.552	No
form1_5	KC 2	low	-1.991	-1.183	No
form1_6	KC 1	low	-24.102	-3.552	No
form1_7	KC 2	low	-24.102	-2.274	No
form1_8	KC 4	low	-24.102	-24.556	Yes
form1_9	KC 7	high	-24.102	-3.552	No
form2_1	KC 5	low	-2.412	-3.752	Yes
form2_10	KC 6	low	-2.412	-1.285	No
form2_11	KC 8	high	-24.023	-3.752	No
form2_12	KC 9	high	-2.412	-3.752	Yes
form2_13	KC 5	low	-2.412	-3.752	Yes
form2_14	KC 9	high	-2.412	-3.752	Yes
form2_15	KC 8	high	-24.023	-3.752	No
form2_16	KC 7	high	-24.023	-25.441	Yes
form2_17	KC 5	low	-1.043	-2.972	Yes
form2_2	KC 5	low	-1.043	-0.059	No
form2_3	KC 3	low	-24.023	-3.752	No
form2_4	KC 8	high	-1.043	-3.752	Yes
form2_5	KC 8	high	-2.412	-25.441	Yes
form2_6	KC 8	high	0.243	-2.972	Yes
form2_7	KC 7	high	-2.412	-4.748	Yes
form2_8	KC 8	high	-2.412	-4.748	Yes
form2_9	KC 9	high	-24.023	-25.441	Yes
form3_1	KC 6	low	-23.291	0.461	No
form3_10	KC 5	low	0.346	3.275	No
form3_11	KC 7	high	-1.604	0.461	No
form3_12	KC 8	high	-2.524	-21.218	Yes

form3_13	KC 9	high	-0.319	1.931	No
form3_2	KC 1	low	-23.291	-21.218	No
form3_3	KC 1	low	-2.524	0.461	No
form3_4	KC 8	high	-0.933	0.461	No
form3_5	KC 1	low	-23.291	-21.218	No
form3_6	KC 2	low	-0.933	24.751	No
form3_7	KC 4	low	-1.604	1.931	No
form3_8	KC 2	low	-1.604	3.275	No
form3_9	KC 1	low	1.255	1.931	No
form4_1	KC 2	low	-0.696	0.131	No
form4_10	KC 9	high	0.069	22.846	No
form4_11	KC 9	high	0.822	22.846	No
form4_12	KC 2	low	-0.696	-22.713	Yes
form4_13	KC 4	low	-1.703	-1.157	No
form4_14	KC 9	high	-0.696	-22.713	Yes
form4_15	KC 8	high	-0.696	-22.713	Yes
form4_16	KC 8	high	-1.703	0.131	No
form4_17	KC 9	high	0.822	0.131	Yes
form4_18	KC 8	high	-1.703	-1.157	No
form4_19	KC 7	high	0.822	0.131	Yes
form4_2	KC 6	low	-1.703	-22.713	Yes
form4_20	KC 6	low	-0.696	-1.157	Yes
form4_21	KC 2	low	0.069	-1.157	Yes
form4_22	KC 8	high	-0.696	1.374	No
form4_23	KC 2	low	0.822	-1.157	Yes
form4_24	KC 8	high	0.822	1.374	No
form4_3	KC 9	high	-1.703	-22.713	Yes
form4_4	KC 5	low	-0.696	0.131	No
form4_5	KC 9	high	0.822	-22.713	Yes
form4_6	KC 6	low	-1.703	-1.157	No
form4_7	KC 9	high	0.069	0.131	No
form4_8	KC 9	high	-0.696	0.131	No
form4_9	KC 8	high	0.822	-1.157	Yes

Knowledge Components by Study

The knowledge components used in each study were selected in collaboration with classroom instructors and designed to include rules that we expected students to have seen before as well as rules that we expected were unfamiliar to the students.

Pilot Study: 8 Knowledge Components

KC 12: Use ‘the’ when the noun has already been mentioned.

KC 13: Use ‘the’ when the noun has already been mentioned.

KC 14: Use ‘the’ when the noun is made definite by a prepositional phrase.

KC 15: Use ‘the’ when the noun is made definite by an adjective clause or an adjective phrase.

KC 2: Use ‘a’ when a singular count noun is indefinite and followed by a consonant sounds.

KC 23: Use ‘the’ with ordinal numbers and ranking words like ‘next’ and ‘last’.

KC 3: Use ‘an’ when a singular count noun is indefinite and the article is followed by a vowel sound.

KC 40: Use ‘the’ for compass directions if they follow prepositions like ‘to’, ‘in’, ‘on’, ‘at’, or ‘from’

Study 1: 12 Knowledge Components

Note: Students were assessed on all 12 KCs but only tutored on 8 in order to assess the validity of the knowledge component model. See Section 4.2 for more information.

KC 2: Use ‘a’ when a singular count noun is indefinite and followed by a consonant sounds.

KC 3: Use ‘an’ when a singular count noun is indefinite and the article is followed by a vowel sound.

KC 5a: Do not use an article with a non-count noun that is indefinite.

KC 5b: Do not use an article with a plural noun that is indefinite.

KC 7: Use ‘a’ or ‘an’ for single letters and numbers.

KC 8/9: Use ‘a’/'a' to mean ‘for each’ or ‘per’ when the noun begins with a consonant/vowel.

KC 12: Use ‘the’ when the noun has already been mentioned.

KC 13: Use ‘the’ when the noun is already known.

KC 14: Use ‘the’ when the noun is made definite by a prepositional phrase.

KC 15: Use ‘the’ when the noun is made definite by an adjective clause or an adjective phrase.

KC 23: Use ‘the’ with ordinal numbers and ranking words like ‘next’ and ‘last’.

KC 50: Use ‘the’ with the word ‘same’.

Studies 2, 3, 4: 6 Knowledge Components

KC 1: Use a or an when a singular count noun is indefinite.

KC 5a: Do not use an article with a non-count noun that is indefinite.
KC 5b: Do not use an article with a plural noun that is indefinite.
KC 7: Use ‘a’ or ‘an’ for single letters and numbers.
KC 12: Use ‘the’ when the noun has already been mentioned.
KC 50: Use ‘the’ with the word ‘same’.

Transfer to Production

One of the long-term goals of second language instruction is to enable students to produce error-free speech and text. While instructing and assessing students using cloze tasks is common classroom practice, educators are also interested in how students perform on more open-ended measures like speeches and essays. Because these studies were run in an in vivo setting, I have access to data beyond the measures collected solely for the study. In particular, I was able to collect writing assignments that were completed as part of normal classroom and homework activities. In total, I collected 220 total essays from 50 students who had participated in Study 2 and/or Study 3. The essays covered a wide variety of topics with sample prompts including: reactions to winning the lottery, describing one's hometown, items one would bring to a desert island, new experiences in Pittsburgh and worst food experiences. With the help of a research assistant, every noun phrase in the 220 essays was coded using a two-step coding process. First each noun was coded for whether it was a proper noun, pronoun, or used another modifier (e.g., *this*, *that*, *some*, *my*, etc.). In the second step, if a noun used an article (*a, an, the* or no article), both the given and correct article were noted. Finally, using the knowledge component model, a code was given to the noun phrase reflecting the feature most important for identifying the correct article. In total, 9,575 noun phrases were coded, and of these 4,106 noun phrases used an article and received a knowledge component code. In addition, each essay was labeled either "before" or "after" to indicate whether the essay was written before or after the article tutor was used in class.

Results

The coding process revealed a large variation in the frequency that students use the various knowledge components. Of the six tutored knowledge components, only four were used frequently, but they accounted for 46% of all the noun phrases that used articles (See Table Below).

Percent correct, incorrect, and representation of Tutored Knowledge Components in Student Writing Samples

Knowledge Component	% Correct	% Incorrect	% of Tutored KCs	% of All KCs
Use *the* if the noun has already been mentioned.	93.8% (n=45)	6.25% (n=3)	2.55%	1.17%
Use *no article* if the noun is general and non-count	82.8% (n=365)	17.2% (n=76)	23.4%	10.7%
Use *no article* if the noun is general and plural.	87.8% (n=423)	12.2% (n=59)	25.6%	11.7%
Use *a* or *an* if the noun is general and singular	71.8% (n=655)	28.2% (n=257)	48.4%	22.2%
Use *the* noun if the noun is modified with the word same.	100% (n=1)	0.00% (n=0)	0.05%	0.02%
Use "a" or "an" if the noun is a letter or number.	0.00% n=0	0.00% n=0	0.00%	0.00%
All Other (non-tutored) KCs	71.8% n=1842	28.2% n=380	n/a	54.1%

Overall, students showed little improvement between essays collected before and after they used the tutor. A mixed model analysis grouping by student and using average correct as the dependent measure and collection time (pre-tutoring vs. post-tutoring) as the dependent factor, reveals no significant difference between the pre-tutoring average (M=78.7%, SD=14.4) and the post-tutoring average (M=81.7%, SD=13.0) ($F(1,79.12)=1.57$, $p=0.21$). I repeated this analysis only including the six knowledge components that were included in the tutor and saw nearly identical results (Tutored KCs Only Pre-Tutor Average = 75.7% SD=21.3, Tutored KCs Only Post-Tutor average = 79.0%, SD=18.8) ($F(1, 78.87)=0.95$, $p=0.33$). Furthermore, there was no effect of condition ($F(3,119.45)=0.64$, $p=0.59$).

Perhaps the most striking result is the relatively high performance on the written assignments. In writing assignments that were collected before tutoring, students used the target knowledge components with 77% accuracy; however, when presented with the cloze-style pretest during the study, students averaged around 69% correct. A paired samples t-test confirms that students did significantly better on writing samples than on the cloze pretest ($t(48)=2.81$, $p=0.007$). This result replicates previous findings that show that students are more accurate on written rather than forced-choice assignments (Butler, 2002; Kharma, 1981; Mizuno, 1985; Tarone, 1985; Tarone & Parrish, 1988). One reason for the performance difference between the two types of tasks could be error avoidance. Error avoidance is a strategy by which students rephrase or restructure their writing in order to avoid using constructs or knowledge components of which they are unsure or not confident.