A Set of Experiments Investigating Methods to Improve Student Learning Through Self-Regulated Learning

Kim M. Kelly
kkelly@wpi.edu
A SET OF EXPERIMENTS INVESTIGATING METHODS TO IMPROVE STUDENT LEARNING THROUGH SELF-REGULATED LEARNING

by

Kim M. Kelly

A Dissertation

Submitted to the Faculty

of

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Learning Sciences and Technology

September 15, 2018

APPROVED:

________________________________________________________________________

Neil T. Heffernan, PhD., Major Advisor

________________________________________________________________________

Ivon Arroyo, PhD.

________________________________________________________________________

Erin R. Ottmar, PhD.

________________________________________________________________________

Ryan S. Baker, PhD., University of Pennsylvania, Graduate School of Education
Abstract

Educators and educational researchers constantly strive to find effective instructional methods that meet the needs of struggling students. There is a well-established relationship between self-regulated learning and academic achievement. Therefore, a great deal of research has been conducted examining the effectiveness of interventions designed to develop self-regulated learning sub-processes including goal setting, help-seeking behavior, self-monitoring, and causal attributions. One particular sub-process that has gained significant attention is self-motivation beliefs, which includes goal orientation. Developing a growth mindset, or the belief that that intelligence is malleable, has been found to increase student learning. Intelligent tutoring systems have also been incorporated into K-12 education to help differentiate instruction and improve learning outcomes. There have been several empirical studies that have attempted to develop help-seeking behavior and growth mindset with interventions delivered by intelligent tutoring systems.

Initially, the goal of this dissertation was to increase student learning by developing self-regulated learning through the use of an intelligent tutoring system. Preliminary attempts failed to modify student beliefs and behavior. As a result, a series of additional randomized controlled trials were conducted. This dissertation is a compilation of those studies, which attempted to leverage ASSISTments, an intelligent tutoring system, to improve student learning in mathematics. Each randomized controlled trial introduced an intervention, based on prior work, designed to address at least one aspect of self-regulated learning and measure the effect on learning. Most of the studies were unsuccessful in producing significant changes in either self-regulation or learning, failing to support the findings of prior research. Survey results suggest that students are reluctant to engage in certain self-regulated learning behaviors, like self-recording, because of the frustration caused when answering a question incorrectly. Based on the findings from these studies, recommendations for potential interventions and future research are discussed.
Acknowledgements

I would like to take this opportunity to acknowledge and thank the many people who have supported me throughout my doctoral studies, without whom this dissertation would not be possible. There are far too many to name here, but I am extremely grateful to everyone who has contributed to my success.

First and foremost, I must thank my husband, Dan and children, Emily, Dylan and Lilly. Their patience and sacrifice over the last seven years did not go unnoticed. Words cannot express my love and gratitude for my husband whose support throughout this journey allowed me to pursue my dream. Attending graduate school while working full-time and raising three young children is not possible without the support of family. Additionally, my in-laws, Marty and Joanne, were always there to help with my children and offer encouraging words. It was my father, Chuck, who urged me to consider a PhD and always enjoyed hearing about my research. My mother, Donna, inspired my love of learning and passion for education. I wish that she were still with us to share this accomplishment with me.

Next I must acknowledge the unwavering support and patience of my advisor, Neil. I will be eternally grateful for this opportunity. His guidance and confidence in me has been greatly appreciated. I would also like to recognize Cristina, who showed me how to balance graduate work and family. I am very grateful for the faculty of the Learning Sciences and Technology program at Worcester Polytechnic Institute, especially the members of my committee who have guided me and challenged me to grow as a researcher. I would also like to acknowledge the support of my peers in the program: Jozsef P., Mike S., Eric V., Seth A., Doug S., and especially Korinn O. and Anthony B., who directly contributed to this dissertation.

Lastly, I must thank all of my students, past and future who inspire me everyday to be a better teacher.
Funding

I would like to thank Professor Heffernan and the grants that supported him and the development of ASSISTments including multiple current NSF grants (ACI-1440753, DRL-1252297, DRL-1109483, DRL-1316736, DGE-1535428, OAC-1724889, OAC-1636782 & DRL-1031398), the US Dept. of Ed (IES R305A180401, R305A120125 & R305C100024 and GAANN), and the ONR.

The research conducted as part of this dissertation was supported by several different funding sources including a Partnership in Math and Science Education (PIMSE) Fellowship (NSF DGE-0742503), a Graduate Assistance in Areas of National Need (GAANN) Fellowship from the U.S. Department of Education (P200A120238), and a Bill and Melinda Foundation Next Generation Learning Challenge via EDUCAUSE.
# Table of Contents

ABSTRACT .................................................................................................................. I

ACKNOWLEDGMENTS .......................................................................................... II

TABLE OF CONTENTS ........................................................................................ IV

LIST OF TABLES ................................................................................................ VI

LIST OF FIGURES ................................................................................................ IX

FOREWARD .............................................................................................................. XII

1  BACKGROUND ....................................................................................................... 1  
   Part 1. Self-Regulated Learning ........................................................................ 1  
      1.1 Overview .................................................................................................. 1  
      1.2 Measuring Self-Regulated Learning ....................................................... 9  
      1.3 Developing Self-Regulation .................................................................. 12
   Part 2. Intelligent Tutoring Systems ................................................................. 15  
      1.4 Overview ................................................................................................ 15  
      1.5 Developing Self-Regulation with Intelligent Tutoring Systems ......... 16

2  THE EFFECT OF SELF-MONITORING ON LEARNING .................................. 26  
   2.1 Background ............................................................................................... 27  
   2.2 Methods ..................................................................................................... 29  
   2.3 Results ........................................................................................................ 32  
   2.4 Conclusions ................................................................................................. 42

3  MOTIVATING THE UNMOTIVATED ............................................................... 47  
   3.1 Introduction ............................................................................................... 48  
   3.2 Study 1 ....................................................................................................... 51  
   3.3 Study 2 ....................................................................................................... 55  
   3.4 Contributions and Discussion .................................................................. 58

4  GROWING A GROWTH MINDSET ................................................................. 64  
   4.1 Background ............................................................................................... 65  
   4.2 Experimental Design ............................................................................... 68  
   4.3 Results .......................................................................................................... 81  
   4.5 Discussion ................................................................................................... 89
# Table of Contents

5  **WHO WANTS TO SELF-MONITOR?** ................................................................. 93
   5.1 Background .......................................................................................... 94
   5.2 Baseline Data Collection ....................................................................... 95
   5.3 Randomized Controlled Trial ................................................................. 96
   5.4 Follow Up Study and Survey ................................................................. 101
   5.5 Discussion and Future Work ................................................................. 105

6  **STOP! PREVENTING HINT OVERUSE TO INCREASE LEARNING** ............ 109
   6.1 Background .......................................................................................... 110
   6.2 Study 1 Methods .................................................................................. 112
   6.3 Study 1 Results .................................................................................... 116
   6.4 Study 2 Methods .................................................................................. 122
   6.5 Study 2 Results .................................................................................... 123
   6.6 Discussion ............................................................................................. 124

7  **ADDRESSING HINT UNDERUSE WITH CHOICE OF AN ACADEMIC TUTORIAL** ................................................................................................................. 129
   7.1 Background .......................................................................................... 130
   7.2 Experimental Design ............................................................................ 132
   7.3 Results .................................................................................................. 137
   7.4 Discussion ............................................................................................. 139

8  **SYNTEHSIS** ............................................................................................ 142

APPENDIX .......................................................................................................... 148
List of Tables

Table 1.1. The sub-processes and definitions of self-motivation beliefs in the forethought phase of self-regulation .................... 5

Table 1.2. Student characteristics of mastery and performance goal orientations (Svinicki, 2005) ................................................................. 6

Table 1.3. Examples of each method or strategy that make up the self-control sub-process of the performance phase in the self-regulated learning model. ............................................................................. 8

Table 1.4. The experimental design used to teach growth mindset in the landmark Blackwell study. (Image taken from Blackwell et al., 2007, page 255.) ................................................................. 13

Table 1.5. A summary of the findings for seven studies that measure the transferability of SRL skills taught through SRL scaffolds. (Image copied from Roll et al., 2014, page 174) ...... 17

Table 2.1. Triplet identification number with a description of the mathematical content and corresponding question numbers from the homework assignment. .................................................................. 30

Table 2.2. Averages, standard deviations (in parenthesis), and effect size for each dependent measure by condition. *Notes a reliable difference. .................................................................................... 33

Table 2.3. The number of students answering all three questions in a triplet incorrectly by triplet identification number and condition ........................................................................................................... 35

Table 2.4. Mean (and standard deviations) learning gains by condition and knowledge level. .................................................................................... 37

Table 2.5. Class time spent (in minutes and seconds) on homework review and the specific homework question numbers covered for each class by condition ........................................................................ 39

Table 3.1. Means and standard deviations (in parenthesis) for survey questions on each task by condition .......................................................... 54

Table 3.2. Mean and standard deviation (in parenthesis) for percent of correct answers on knowledge tests ..................................................... 54

Table 3.3. Means and standard deviations (in parenthesis) of survey questions on each task by condition and for percent of correct answers on knowledge tests ........................................................................ 57
Table 4.1. Text messages embedded in the assignment to reinforce specific attributes depending on condition and performance.77

Table 4.2. The experimental design by condition. Shared assignments and differences between conditions are highlighted. .............79

Table 4.3. Student self-belief survey results by condition for pre- and post- intervention compared to the norm-referenced values. *Indicates a significant difference. ..........................................83

Table 4.4. Mean (and standard deviations in parenthesis) levels of frustration reported in general and during the target learning assignment by condition using a scale of 1 to 4 where lower values indicate higher frustration. .......................86

Table 4.5. Changes in mean learning gains between baseline and post-test by condition .................................................................88

Table 4.6. The number of students (and percentages) by condition whose learning gains between baseline and post-test “gained”, “lost”, or “stayed the same”.... .................................................................88

Table 5.1. Student responses, by condition, when asked if they would like to receive feedback during two different assignments. 100

Table 6.1. Study design showing timing and function of each experimental element. .............................................................................114

Table 6.2. Means and standard deviations in parenthesis for baseline assignment by condition including performance as a percent, overall hint count, bottom-out hint count and learning as a percent increase.................................................................116

Table 6.3. Means and standard deviations in parenthesis for intervention assignment by condition including performance as a percent, overall hint count, bottom-out hint count and learning as a percent increase. .........................................................118

Table 6.4. Means and standard deviations in parenthesis for baseline assignment, for target population only, by condition including performance as a percent, overall hint count, bottom-out hint count and learning as a percent increase. .........................................................119
Table 6.5. Means and standard deviations in parenthesis for Intervention Assignment, for target population only, by condition including performance as a percent, overall hint count, bottom-out hint count and learning as a percent increase. ..................................................................................................................................................120

Table 6.6. Summary of shifts in bottom-out hint count and learning gains between baseline assignment and intervention assignment for target population only. .................................................121

Table 6.7. Average bottom-out hint count and learning gains by condition.............................................................................................................................................................................................................124

Table 7.1. The number of students in the treatment condition choosing to receive the tutorial when asked initially (first question choice) and after making an error (second question choice). ..................................................................................................................................................138

Table 7.2. The number (and percentage) of students who completed the assignment by condition...................................................................................................................................................................................139
List of Figures

Figure 1.1. Cyclical model of self-regulated learning highlighting the three phases (forethought, performance, and self-reflection) and the sub-processes within each phase (Zimmerman & Campillo, 2003)................................................. 4

Figure 2.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, metacognitive self-monitoring, is highlighted. ................................................................. 26

Figure 2.2. The feedback provided by ASSISTments, as viewed by students in the WBH (feedback) condition, following a correct response. ................................................................. 31

Figure 2.3. The feedback provided by ASSISTments, as viewed by students in the WBH (feedback) condition, following an incorrect response. ................................................................. 31

Figure 2.4. The feedback provided by ASSISTments, as viewed by students in the WBH (feedback) condition, when a hint was requested. ................................................................. 31

Figure 2.5. The feedback provided by ASSISTments to students in the TH (no feedback) condition after submitting an answer. ................................................................. 32

Figure 3.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, self-motivational beliefs, specifically outcome expectations, and task value/interest, are highlighted. ................................................................. 47

Figure 4.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The present research focuses on addressing several aspects of this cycle while evaluating the impacts on others. Specifically, the intervention attempts to shift students’ self-motivation beliefs and causal attribution. The impact on help seeking behaviors and learning were measured. ................................................................. 64

Figure 4.2. Logic model for experimental design................................................. 67
Figure 4.3. Sample question from Baseline Learning Task showing the different forms of help available upon request. ................................................................. 70

Figure 4.4. A question in the final intervention assignment designed to assess growth mindset in students. ................................................. 74

Figure 5.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, metacognitive self-monitoring, is highlighted................................................................. 93

Figure 5.2. Instructions at the start of a homework assignment that presents students with the choice to receive correctness feedback or not. ................................................................. 95

Figure 5.3. The message provided to students who chose not to receive feedback when entering their response to a question on the homework assignment. ................................................................. 97

Figure 5.4. The feedback message provided to students who selected feedback, after entering a correct response........................................... 97

Figure 5.5. The feedback message provided to students who selected feedback, after entering an incorrect response............................. 97

Figure 5.6. The initial question in the homework assignment for students in the experimental condition. It includes the intervention, a video, and offers students a choice regarding whether they want to receive feedback or not......................... 98

Figure 5.7. The first question of the homework assignment which provides a choice to students whether they would like to receive feedback or not. ................................................................. 101

Figure 5.8. The intervention, a text message, shown to students in the experimental condition who initially requested not to receive feedback designed to encourage students to select feedback. .................................................................. 102

Figure 5.9. The question immediately following the intervention providing students another opportunity to request feedback. .................................................................. 102
**Figure 5.10. Sample student responses (intentionally unedited) when asked to explain why they requested correctness-feedback or not.**

**Figure 6.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, help seeking behavior, is highlighted.**

**Fig. 6.2. Sample hints available upon request to students.**

**Fig. 6.3. The intervention.**

**Figure 6.4. Directions presented to students explaining the remaining portion of the assignment.**

**Figure 7.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, help seeking behavior, is highlighted.**

**Figure 7.2. The initial question for students in the treatment condition that offered students a choice to complete a tutorial or not before beginning the assignment.**

**Figure 7.3. The question posed to students in the treatment condition who initially refused the tutorial but proceeded to incorrectly answer the first question in the assignment.**

**Figure 7.4. The experimental designs by condition.**

---

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.10</td>
<td>Sample student responses (intentionally unedited) when asked to explain why they requested correctness-feedback or not.</td>
<td>104</td>
</tr>
<tr>
<td>6.1</td>
<td>The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman &amp; Campillo, 2003). The focus of this research, help seeking behavior, is highlighted.</td>
<td>109</td>
</tr>
<tr>
<td>6.2</td>
<td>Sample hints available upon request to students.</td>
<td>114</td>
</tr>
<tr>
<td>6.3</td>
<td>The intervention.</td>
<td>115</td>
</tr>
<tr>
<td>6.4</td>
<td>Directions presented to students explaining the remaining portion of the assignment.</td>
<td>123</td>
</tr>
<tr>
<td>7.1</td>
<td>The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman &amp; Campillo, 2003). The focus of this research, help seeking behavior, is highlighted.</td>
<td>130</td>
</tr>
<tr>
<td>7.2</td>
<td>The initial question for students in the treatment condition that offered students a choice to complete a tutorial or not before beginning the assignment.</td>
<td>134</td>
</tr>
<tr>
<td>7.3</td>
<td>The question posed to students in the treatment condition who initially refused the tutorial but proceeded to incorrectly answer the first question in the assignment.</td>
<td>135</td>
</tr>
<tr>
<td>7.4</td>
<td>The experimental designs by condition.</td>
<td>137</td>
</tr>
</tbody>
</table>
FOREWORD

The field of education has become inundated with technology as schools across America provide access to iPads, Chromebooks, or other personal technology. A variety of software and programs such as Google classroom, apps, games, and a plethora of intelligent tutoring systems are commonplace in K-12 instruction. On a weekly basis, teachers are bombarded with ads for the latest program that promises to provide customized practice to students in order to enhance learning. Individualized instruction is alluring in a time when differentiation is paramount yet challenging. However the basis for all of these programs is individual learning, which requires students to be self-regulated learners. That is to say, they must know how to learn and possess the skills necessary to learn independently. Of course, intelligent tutoring systems attempt to design features to compensate for students’ lacking self-regulatory skills. But are they effective? Can technology actually replace lacking learning strategies? If so, can they simultaneously develop these strategies as students practice topics, such as math, so that students become better self-regulated learners?

These questions were the initial driving force for this dissertation. However, what makes this research unique is that I am a practicing 7th grade math teacher. As such, the research is much more than a culmination of randomized controlled trials that I conducted over the last seven years. Each of the presented studies was conducted in an attempt to address an observed problem of practice and together they tell a story and shed light on some of the challenges in K-12 education today.

The story begins with my introduction to ASSISTments, an intelligent tutoring system, designed to support math instruction. So often in education we are presented with “research-based” practices and tools without any evidence of their effectiveness. As a researcher, I was able to conduct a randomized controlled trial testing the efficacy of correctness feedback on nightly homework (Chapter 2). The results confirmed what I already knew, learning gains were incredible as a result of immediate
correctness feedback. However, that is not the end of the story. I observed that these results were limited to my honors classes. My general education classes did not appear to experience the same learning gains. Hypothesizing that motivation was in part responsible for the differences, I conducted another randomized controlled trial focusing on motivating my unmotivated students. While the learning gains were not astonishing, the research suggested that it was possible to shift student beliefs through a fairly simple intervention (Chapter 3), which over time may impact learning.

About this time, Carol Dweck’s growth mindset work was becoming popular in K-12 education. I connected mindset to self-regulation and hypothesized that if I could shift students’ mindsets while simultaneously teaching students how to learn from their mistakes that I could develop necessary self-regulation skills all while learning math content (Chapter 4). This goal proved too lofty. Shifting behavior was dependent on shifting mindset, which, results indicate, is very challenging potentially for multiple reasons.

As a result, I attempted to focus on developing one specific self-regulated behavior, requesting feedback as a form of self-monitoring (Chapter 5). Despite two different attempts at encouraging students to request feedback, I failed to shift this behavior. Narrowing my focus still, I chose to focus on the students who did request feedback yet failed to use that feedback effectively to learn (Chapter 6). Students who request all available hints and then simply request the answer do not seem to learn. I hypothesized that if I could encourage students to use the hints, but stop before requesting the final answer and force them to read the provided hints that learning would increase. Initial results suggested this was a promising intervention, however a replication study failed to produce the same results.

In a final attempt to help my students learn more effectively, and to develop self-regulation strategies, I conducted a randomized controlled trial that provided students the option to receive a tutorial before beginning a specific assignment. In general, students were willing to participate in the tutorial, especially after learning they
answered a question incorrectly. However, providing students a choice as well as allowing students to use the hints, penalty free, did not seem to increase learning.

I initially set out to develop my students’ self-regulation strategies so that they would become more effective learners. Despite promising, successful interventions in the literature, I failed in almost every attempt. I then tried to force self-regulated strategies on my students hoping to see improved learning. This too, was unsuccessful. This suggests that shifting mindset and behavior will remain an ongoing challenge for educators. Additionally, for some students with insufficient self-regulatory strategies, on-line and independent learning must be supplemented with teacher intervention. Future work should continue to attempt to address the needs of struggling students to compensate for lacking self-regulation strategies as well as to teach those strategies.
1 Background

For decades, a teacher’s job has been to teach students content. Thanks to the technological revolution, content is readily available on the Internet, and as a result, the role of the teacher has shifted. Teachers must now teach students how to learn so that they may access the breadth of information available to them. Self-regulated learning refers to the skills, strategies, and processes necessary to learn independently. Being a strong self-regulated learner is imperative as more and more learning is being done independently through the use of technology.

This chapter is divided into two sections. The first section will provide an overview of Self-Regulated Learning (SRL) including the tools used to measure SRL and a review of the literature that suggests how to develop SRL in students. The second section will explore Intelligent Tutoring Systems (ITS), highlighting features of current ITS that attempt to simulate self-regulation strategies for students as well as interventions designed to develop those self-regulation strategies.

Part 1. Self Regulated Learning

1.1 Overview

In the field of psychology, self-regulation refers to the ability to control one’s own behavior, emotions, and thoughts. Self-Regulation Theory (SRT) has been widely studied (see Boekaerts, Pintrich & Zeidner, 2000 for an overview). The theory suggests that self-regulation is a “system of conscious personal management that involves the process of guiding one's own thoughts, behaviors, and feelings to reach goals” (Baumeister, Schmeichel & Vohs, 1996). Baumeister, a social psychologist, suggests that there are four components that interact to determine our behavior. They include standards of desirable behaviors, motivation to meet standards, monitoring of situations and thoughts, and willpower to control impulses (Baumeister, Heatherton & Tice, 1994). Essentially, SRT helps explain why we get up and go to
work when we would rather sleep all day, or why we do not eat a gallon of ice-cream while on a diet.

Expanding on SRT, the Social Cognitive Theory of Self-Regulation was developed. Instead of examining self-regulation through Baumeister's four components, Bandura (1991) describes self-regulation as the product of three sub-functions: *self-monitoring, judgment, and self-reaction*. These sub-functions incorporate and expand the components of SRT. *Self-monitoring* refers to the process of setting goals and monitoring one’s progress towards reaching those goals. *Judgment* occurs when one compares their goals, performance, or behavior to socially acceptable norms and/or their own values. *Self-reaction* refers to the affect assigned based on the judgment. For example, people may feel pride or disappointment following an action. An anticipated self-reaction may then impact goals and the decision-making progress, creating a feedback loop.

The Social Cognitive Theory of Self-Regulation also introduces *self-efficacy*, an additional mechanism of self-regulation. *Self-Efficacy* refers to one’s personal judgment of “how well one can execute courses of action required to deal with prospective situations” (Bandura 1982). Self-beliefs of efficacy, or ability, heavily impact each of the sub-functions included in the Social Cognitive Theory. Specifically, self-efficacy impacts goal setting. If one believes one can achieve something, one tends to set high goals for themselves and persevere despite challenges.

How do the Self Regulation Theory and Social Cognitive Theory of Self-Regulation apply to learning processes? To answer this question, let us consider the following two students. Dan and Emily are both struggling to complete a math homework assignment. Dan believes that he is not good at math (self-efficacy) so when he isn’t sure how to solve the problem he thinks he is a failure (self-reaction) and quits (motivation). Emily, on the other hand, thinks she’s good at math (self-efficacy). When she is stuck on a problem, she looks back in her notes to find a similar problem. That was not particularly helpful (self-monitoring) so she watches a video on
the topic. She perseveres and is finally able to complete the assignment and is proud of herself (self-reaction). Dan is unable to effectively guide his own behaviors in order to learn, whereas Emily takes control of her learning experience and as a result learns the math. While components of SRT and Social Cognitive Theory of Self-Regulation are easily identifiable in these situations, learning is so complex that a direct mapping of self-regulation to learning is helpful.

Self-Regulated Learning (SRL) applies self-regulation theories to the complex domain of learning. SRL is defined as the degree to which students are metacognitive, motivated, and behaviorally active participants in their own learning process (Zimmerman, 1986). Over the years, researchers have developed multiple models of SRL that attempt to represent the interactions between the cognitive, metacognitive, behavioral, motivational, and emotional/affective aspects of learning (Zimmerman & Campillo, 2003; Boekaerts, 1991, Boekaerts & Cascallar, 2006, Winne & Hadwin, 1998, Pintrich, 2000, Efklides, 2011, Jarvela & Hadwin, 2013). Panadero (2017) provides a thorough review and comparison of these models. Interestingly, all of the models agree that SRL is cyclical with multiple sub-processes that can be organized into at least three phases; (1) preparatory, which includes task analysis, planning, and goal setting, (2) performance of the task, which includes monitoring of progress, and (3) appraisal, which includes reflection, regulation, and adaptation for future performances.

The most widely cited model is Zimmerman (2000) (Panadero, 2017), as such, the adapted model that includes the sub-processes (Zimmerman & Campillo, 2003), is the model that will be used as the basis for this entire dissertation (see Figure 1.1). This model defines the three phases of SRL as the forethought phase, the performance phase, and the self-reflection phase. Within each phase, sub-processes related to motivation, metacognition, and behavior are included.
Forethought phase:
According to this model, prior to starting a task, students pass through the forethought phase. This phase consists of two major classes of sub-processes, task analysis and self-motivation beliefs. Task analysis includes goal setting and strategic planning. Goal setting is a topic with a vast body of its own research, which falls outside the scope of this work. However, one common finding that is worth mentioning here is that setting long-term goals with proximal sub-goals leads to increased learning (Weldon, 1998). Let us consider the following example to better understand these sub-processes. A student in an elementary math class may set a goal to learn her multiplication facts by the end of the year. She may establish a sub-goal of learning one factor a week. This demonstrates both goal setting and strategic planning.

Figure 1.1. Cyclical model of self-regulated learning highlighting the three phases (forethought, performance, and self-reflection) and the sub-processes within each phase (Zimmerman & Campillo, 2003).
The second major class of sub-processes in the forethought phase is known as self-motivation beliefs, which can be used to describe students’ beliefs about their learning and have a significant impact on motivation. These beliefs include outcome expectations, task value/interest, goal orientation, and self-efficacy. Each of these sub-processes can be individually defined, but are heavily dependent on one another. Table 1.1 defines each sub-process.

Table 1.1. The sub-processes and definitions of self-motivation beliefs in the forethought phase of self-regulation.

<table>
<thead>
<tr>
<th>Sub-Process</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome Expectation</td>
<td>Perceived purpose of the task</td>
</tr>
<tr>
<td>Task Value/Interest</td>
<td>Perceived value of the task</td>
</tr>
<tr>
<td>Goal Orientation</td>
<td>Perceived value of the learning process</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>Perceived ability to learn</td>
</tr>
</tbody>
</table>

Outcome expectations refer to the perceived purpose of the task and task interest refers to the perceived value of the task, which can be influenced by the purpose of that task. The relationship between outcome expectation, task value and motivation appears intuitive; if an outcome or task is important enough we are highly motivated to complete that. Pekrun (2006) empirically showed that increased task value leads to increased motivation. Let us continue to consider the example of the elementary student learning her multiplication facts. She perceives the purpose of the task is to prepare her for learning more math in the future, which is important to her; she believes that learning her facts now will make math easier later on. These beliefs help her to be highly motivated to learn her multiplication facts.

Goal orientation refers to the value a student places on the learning process. A strong goal orientation means that a student values the learning process because they value learning new things. The Achievement Goal Orientation Theory provides labels and descriptions for the opposing orientations, “mastery goal orientation” and
“performance goal orientation” (Dweck & Leggitt 1988). Consider these two students. Dylan loves to learn math and is excited to learn his multiplication facts because it is a fun new challenge. Lilly is only learning her math facts because she does not want to fail the quiz. According to the achievement goal orientation theory, Dylan has a mastery goal orientation because his goal is to learn new content. Lilly has a performance goal orientation because her goal is based on a grade or at least to avoid performing poorly. Table 1.2 highlights the differences between these goal orientations.

Table 1.2. Student characteristics of mastery and performance goal orientations (Svinicki, 2005).

<table>
<thead>
<tr>
<th>Mastery Oriented Students</th>
<th>Performance Oriented Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus on skill/content</td>
<td>Focus on appearing competent or favorable comparisons to peers</td>
</tr>
<tr>
<td>Enjoys challenges</td>
<td>Avoids challenges, prefers comfortable tasks</td>
</tr>
<tr>
<td>Views mistakes as learning opportunities</td>
<td>Views mistakes as evidence of lack of competence</td>
</tr>
</tbody>
</table>

Goal orientation stems from students’ beliefs about intelligence. Students with a mastery goal orientation tend to believe that intelligence is malleable and can be incrementally modified. Students with a performance-based orientation tend to believe that intelligence, or mental ability is a fixed entity. Carol Dweck coined the terms, “growth mindset” and “fixed mindset” to describe these opposing beliefs (Dweck & Leggitt, 1988). Students with a growth mindset tend to set learning goals, whereas students with a fixed mindset tend to set goals that validate their perceived intelligence (Blackwell et al., 2007). Research suggests that, especially in math and science, students with a growth mindset persevere more and outperform students with a fixed mindset (Dweck, 2008).

Self-efficacy refers to one's belief that they are capable of learning. Self-efficacy affects how people feel, think, and motivate themselves (Bandora, 1994). This tends
to be correlated with mindset. Let us consider two different students. Joanne and Martin have both struggled with math in the past. Joanne has a growth mindset, so even though she has not had much success in the past, she has high self-efficacy because she believes that she is capable of learning. Martin has a fixed mindset, so he views his past failures as evidence that he is not good at math and will never be good at math, resulting in low self-efficacy. It is not surprising that self-efficacy also impacts goal setting and motivation. If one believes they will be successful, they are motivated to persist through challenges.

Performance Phase:
The second phase of the SRL model begins when the task itself begins. This phase also has two major classes of sub-process, self-control and self-observation. Self-control refers to the ability to use a variety of specific methods or strategies for learning, such as: imagery, self-instruction, time management, help-seeking, and self-consequences. These strategies and methods are concrete things students can do while engaged in a learning task to ensure learning is successful. Table 1.3 provides an example for each method. These strategies are task and learner dependent (Zimmerman & Moylan, 2009). That means that a student may choose to employ different strategies for different tasks and similarly, different students may employ different strategies for the same task.

The second major class of the performance phase is self-observation. This is essential in analyzing and adjusting the strategies discussed above. Self-observation requires students to both self-monitor and self-record. Self-monitoring is a metacognitive process that entails informally tracking progress and comparing it to one’s goals. For example, a student may notice that he usually gets overwhelmed when reading a novel, but when he creates a schedule and break up the reading he is more successful. Self-recording is a formal process of tracking progress. For example, a student may graph the number of multiplication problems she can answer correctly in two minutes every day for a week in order to track progress.
Table 1.3. Examples of each method or strategy that make up the self-control sub-process of the performance phase in the self-regulated learning model.

<table>
<thead>
<tr>
<th>Self Control Method or Strategy</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imagery</td>
<td>Associating a picture with a new vocabulary word</td>
</tr>
<tr>
<td>Self-Instruction</td>
<td>Quizzing oneself after each section of reading in a textbook</td>
</tr>
<tr>
<td>Time Management</td>
<td>Setting a schedule to complete tasks (read one chapter every day to finish the book on time)</td>
</tr>
<tr>
<td>Help-Seeking</td>
<td>Watching a video on YouTube to help learn how to solve a problem.</td>
</tr>
<tr>
<td>Self-Consequences</td>
<td>Setting a reward for finishing 10 problems</td>
</tr>
</tbody>
</table>

**Self-Reflection Phase:**

The third and final stage of SRL is known as the self-reflection phase and it occurs at the conclusion of the learning task. This phase also consists of two major classes of sub-processes, **self-judgment** and **self-reaction**. The function of **self-judgment** is to evaluate the learning process, which requires students to **self-evaluate** and develop a **causal attribution**. **Self-evaluation** refers to a comparison of oneself against a standard. That standard could be their goal, prior performance, mastery, or social comparison to the performance of others (i.e. grades) (Bandura, 1986).

**Causal attribution** refers to the beliefs about the causes of one’s successes or failures. This attribution then impacts self-efficacy and future motivation (Zimmerman, 2002). For example, a student with a fixed mindset, who attributes his failure to his lack of ability will not have much self-efficacy and therefore lack motivation in a future learning task. Conversely, a student with a growth mindset, who attributes her failure to a poorly chosen learning strategy will be motivated on the next learning task and will be eager to engage a new strategy.
**Self-reaction** takes on two forms, *self-satisfaction* and *defensive reactions*. Completing a learning task and self-evaluation will result in some affective reactions, which can either be positive or negative, regardless of performance. For example, a student who ultimately failed an assignment but worked hard and persevered despite challenges may feel self-satisfaction in the form of pride. A student who completed a challenging assignment but did not score as well as others may have a defensive reaction and attempt to protect one’s self image by appearing apathetic. Students’ self-reactions often depend on their self-evaluation and self-motivation beliefs. Furthermore, those self-reactions then influence future self-motivation beliefs and goal setting generating the cyclical structure of SRL (Zimmerman & Moylan, 2009).

The interdependence of many of the sub-processes and cyclical nature of self-regulated learning creates a snowball effect where self-reflection from one learning task feeds into and influences the forethought phase of the next task, which has significant implications for the outcome of that task, which then provides further evidence for self-reflection and so on and so on. This inertia can be beneficial if learning experiences are positive. For example, a student who has a successful learning experience and feels self-satisfaction will approach the next learning activity with higher self-efficacy and stronger goal orientation (Dweck & Leggitt, 1988), greater task value (Zimmerman & Kitsantas, 1997), and therefore motivation. This will lead to greater success, which during self-reflection will lead to self-satisfaction and so the cycle repeats. Unfortunately, the converse is also true for students who do not have a successful learning experience. They approach the forethought phase with low self-efficacy and lack motivation leading to future poor performances and defensive reactions, which only serve as evidence that they are incapable of learning and the cycle repeats (Zimmerman & Moylan, 2009).

1.2 Measuring Self-Regulated Learning

Self-Regulated Learning is complex because it has so many components that interact with one another and encompasses thoughts, feelings, and behaviors before,
during and after a learning activity. This makes measuring SRL challenging. Nonetheless, researchers have developed a variety of tools to measure SRL that have been found to correlate with performance (Pintrich et al., 1993) and predict academic grades (Zimmerman & Martinez-Pons, 1988). Traditional tools include questionnaires while more recent tools include computer traces and microanalytical measures (Zimmerman, 2008). This section will review several of these tools and discuss the implications of their use in SRL research. For a more thorough review of the instrument and measurement methods see Panadero (2017).

One of the first tools created was the Learning and Study Strategies Inventory (LASSI; Weinstein, Schulte & Palmer, 1987). This questionnaire consists of 80 questions broken into 10 scales that assess skill, motivation, and self-regulation strategies. Another widely used questionnaire is the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1993). This questionnaire consists of 81 questions broken into two major sections: learning strategies and motivation. These major sections are comprised of several different scales that include value, self-efficacy, and anxiety. Both of these questionnaires ask students to respond on a Likert Scale reflecting how they typically feel or in general how true that statement is of them. These tools are considered by some to be a measure of self-regulated learning aptitude in that they measure an aggregate of skills, beliefs and feelings rather than those specific to a particular task (Winne & Perry, 2000). This means they may not be sensitive to fluctuations that may occur within the context of research.

To address this, the On-Line Motivation Questionnaire (OMQ) was created (Boekaerts, 1999). This tool allows us to measure student’s motivation during a specific task or situation in a classroom context. Just prior to starting a task, students respond to a series of questions using a Likert Scale to assess the amount of effort students expect to expend on the task. Immediately following task completion, students report their feelings and attributions. This tool has been found to be valid and reliable for use in research purposes (Crombach, Boekaerts, & Voeten, 2003),
but it should be noted that the SRL sub-processes that are measured with this tool are limited.

More recently, especially with the introduction of online learning, researchers have been designing measurement tools beyond questionnaires to assess SRL. In the context of reading to learn new information, it is possible to trace students’ learning strategies when using the software program gStudy (Winne et al., 2006). While reading a text, the program allows students to engage in a variety of self-regulated behaviors such as: note taking, annotating, creating glossaries, constructing concept maps, even collaborating with peers or tutors. These behaviors can be tracked and recorded allowing researchers to construct a measure of self-regulation and to link those measures to academic outcomes. Using behavior data has been found to be more accurate than self-reports (Winne & Jamieson-Noel, 2002). However, for increased accuracy, trace data should always be combined with other measures of SRL (Zimmerman, 2008).

Micro analytical measures are another measurement tool that can be used when researching SRL, especially in the context of on-line learning. Similar to the OMQ, this tool measures self-regulatory processes and motivation during a specific learning task instead of providing a general measure of self-regulatory behaviors. This is done by asking students open- and closed-ended questions at key moments throughout an assignment, including before, during, and after learning (Zimmerman, 2008). This allows researchers to track self-efficacy across several learning tasks as well as to observe changes in sub-processes within and across the three SRL phases (Zimmerman, 2000). The downside to this measurement tool is that creating the questions, which must be tied directly to the content of the task, and logging and analyzing trace data may be cumbersome.
1.3 Developing Self-Regulation

Strong self-regulated learning skills have been correlated with academic success (Zimmerman & Schunk, 2001). Therefore it seems natural that researchers would attempt to find ways to help struggling students foster SRL skills in order to improve learning outcomes. Several meta-analyses have concluded that it is possible to promote self-regulated learning in students through classroom based interventions focused on teaching learning strategies as well as metacognition (Hattie et al., 1996; Dignath & Buttner, 2008; Dignath, Buttner, & Langfeldt, 2008). This section will highlight a small selection of the research that has been conducted on developing SRL through non-technologically based interventions (see Part 2 of this chapter for technological interventions). I have chosen to focus on those studies with interventions that have successfully improved learning in math or whose interventions have inspired my research.

The SRL model makes it clear that self-efficacy plays an integral role in learning because it affects every phase of self-regulation. In fact, low self-efficacy has been found to have harmful effects on motivation and performance (Bandura & Locke, 2003). An analysis of high school students revealed that the source of academic self-efficacy is prior successes or failures in school (Bryant, 2017). This suggests that once a student develops low self-efficacy, changing that becomes very challenging. Current literature offers suggestions on how to improve self-efficacy based on theory, but does not offer empirical evidence of successful interventions. For example, Margolis & McCabe (2006) recommend that teachers use enactive mastery (using a “small win” as a catalyst for future beliefs), vicarious experiences (using someone else’s experience as an example), and verbal persuasion (positive verbal feedback) in order to strengthen the beliefs of struggling learners.

In addition to self-efficacy, goal orientation also impacts multiple phases of SRL and therefore learning (Grant & Dweck, 2003). Goal orientation, and mindset have been found to be positively correlated (Dweck, 2008; Leggitt & Dweck, 1988; Hong et al.,
Several studies have shown that interventions designed to create a growth mindset have successfully improved learning (Yeager & Walton, 2011). One such study is well known in K-12 education (Blackwell, Trzesniewski, & Dweck, 2007). The intervention consists of eight 25-minute lessons (see Table 1.4 for the experimental design). Students in both the experimental and control conditions were presented lessons on the physiology of the brain, study skills, and nonstereotypic thinking. Students in the experimental condition were also exposed to the growth mindset by teaching them that intelligence is malleable and can be developed. Results indicated that the intervention successfully helped students adopt a growth mindset, increase motivation, and increase academic performance.

Another study found similar results, finding that it is possible to help students develop a growth mindset by teaching about the malleability of the brain (Aronson, Fried & Good, 2002). The intervention used in this study consisted of three one-hour sessions where college students in the experimental condition wrote letters to middle-school pen pals. Participants were encouraged to teach their pen pals about the malleability of the brain, which required the participants themselves to hear the important message that intelligence can be expanded with effort. Again, this shift in mindset leads to increased academic performance.

### Table 1.4. The experimental design used to teach growth mindset in the landmark Blackwell study. (Image taken from Blackwell et al., 2007, page 255.)

<table>
<thead>
<tr>
<th>Sessions</th>
<th>Experimental group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 and 2</td>
<td>The Brain—Structure &amp; Function: Brain Anatomy, Localization of Function, Neuronal Structure, Neurotransmission</td>
<td>Same as experimental group</td>
</tr>
<tr>
<td>3 and 4</td>
<td>Incremental Theory Intervention Reading (aloud in class): “You Can Grow Your Intelligence” Activity: “Neural Network Maze,” showing how learning makes your brain smarter</td>
<td>Alternative Lesson: Memory Reading (aloud in class): “Memory” Activity: “Grocery Store Tricks,” teaching mnemonic strategies</td>
</tr>
<tr>
<td>5 and 6</td>
<td>Anti-Stereotyping Lesson: Slides, activity, discussion to illustrate the pitfalls of stereotyping. Study Skills Lesson: Slides, lecture, discussion, handouts teach time management and study skills.</td>
<td>Same as experimental group</td>
</tr>
<tr>
<td>7 and 8</td>
<td>Discussions: Learning makes you smarter; Labels (e.g., stupid, dumb) should be avoided</td>
<td>Discussions: Academic difficulties and successes, preferences; Memory and the brain</td>
</tr>
</tbody>
</table>
While the interventions presented thus far have been successful, they are time consuming and implementing them at scale would be problematic. Paunesku et al. (2015) created an intervention that attempted to condense and incorporate the interventions from Blackwell et al. (2007) and Aronson et al., (2002). Instead of an eight-week course, the information about growth mindset was condensed into a single article. In place of the pen pal writing assignment, participants completed two short writing tasks. The first asked participants to summarize the scientific information from the article while the second asked participants to offer advice, based on what they read, to a hypothetical student who was struggling and growing frustrated. The entire intervention was completed in a 45-minute session. Results indicated that students developed a more malleable view of intelligence and had significant increases in GPA. These results are encouraging in that this intervention can be used at scale easily across content areas and grade levels.

In addition to mindset, self-reflection is a fundamental aspect of the cyclical nature of SRL in that a student’s self-reflection will directly impact future behavior on future tasks (Weiner, 1986). Therefore encouraging healthy self-reflection may be beneficial in developing SRL. Zimmerman and colleagues (Zimmerman, Moylan, Hudesman, White, & Flugman, 2008) used self-reflection training in order to help high school students in a remedial math class more accurately monitor their mathematical problem solving performance, which lead to higher achievement. The intervention entailed students making self-efficacy judgments before and self-evaluative judgments after quiz questions. After receiving feedback from the teacher, students were then required to complete a self-reflection form. This form asked students to compare their judgments to their performance, explain their ineffectual strategies, develop more effective strategies, and assess their confidence for solving new problems. Forcing students to use performance feedback to self-evaluate helped students to realize that scores are not the end of learning, but instead are used as part of the learning process.
Part 2. Intelligent Tutoring Systems

1.4 Overview

In 1984, Benjamin Bloom detected the educational phenomenon known as the 2-sigma problem. Students receiving one-on-one instruction with a human tutor score two standard deviations higher than students learning with traditional instructional methods (Bloom, 1984). However, it is unrealistic to consider that all instruction will be conducted in such a costly manner. Therefore, in an attempt to mimic the learning gains of a human tutor, developers of Intelligent Tutoring Systems (ITS) have leveraged a variety of features to enhance learning by individualizing the feedback and the pace (Corbett, 2001). These features may include tutoring, which often involves scaffolding. Scaffolding may generate sub-goals for students by outlining the solution process, provide worked examples or models of the expert steps required to solve a problem (Salden et al., 2009), automatize learning tools such as system-triggered hints, or provide correctness feedback at the answer- or step-level (Corbett & Anderson, 2001). Essentially, ITS are computer-based learning systems that attempt to automatically identify the needs of the learner and adapt accordingly providing a customized learning experience.

Numerous studies have shown that ITS do in fact enhance learning (Kulik & Fletcher, 2016; Anderson, et al., 1995; Koedinger et al., 1997). Of course there is variation in the effectiveness of these programs based on the granularity of support (answer-based, step-based, sub-step based). In order to measure “effectiveness” researchers often compare effect sizes, with Bloom’s 2-sigma effect ($d = 2.0$). Computer Aided Instruction (CAI), or programs that provide feedback and support at the answer-level, have been found to have an average effect size of $d = 0.31$ (Kulik & Kulik, 1991). Intelligent Tutoring Systems (ITS), or programs that provide step- or sub-step- level feedback tend to have a larger effect size, $d = 1.0$ (Anderson et al., 1995). VanLehn (2011) found that ITS actually have an effect size of $d = 0.76$. However, he also found that the effect size of human tutors is much lower than Bloom’s two standard
deviation, in fact it is closer to $d = 0.79$. This means that ITS are almost as effective in supporting instruction as human tutors.

While there are many examples of ITS that are currently being used in K-12 education, ASSISTments is the platform that will be used to conduct the research presented throughout the upcoming chapters (Heffernan & Heffernan, 2014).

### 1.5 Developing Self-Regulation with Intelligent Tutoring Systems

We already know that improved SRL skills are correlated with improved academic outcomes. Aleven and Colleagues (2006) have found that students with better self-regulation also achieve better learning outcomes in ITS. One benefit of computer-based learning is that technology can compensate for some of the students’ lacking SRL skills (Winne & Nesbit, 2009). For example, Khan Academy provides users immediate correctness feedback while practicing math, which mimics the self-monitoring sub-process of SRL (Khan, 2018). Other, more complex systems, adapt content or instruction based on learner characteristics (Shute & Zapata-Rivera, 2012). Some ITS developers have embedded interventions for SRL into their programs so that students can develop their SRL strategies while learning content (Azevedo & Aleven, 2013). This section provides a summary of some of the interventions designed to do just that.

Similar to the domain-specific feedback described above, ITS can deliver SRL feedback, or feedback on students’ actions within the system. For example, the Help Tutor is an automatic agent that provides feedback to students regarding their choice to seek or avoid help (Roll et al., 2011). Researchers first sought to identify poor help-seeking behavior and then provided messages to students who engaged in that behavior. An example of a message given to students who were requesting hints faster than they could use them is, “No need to hurry so much. Take your time and read the hint carefully. Consider trying to solve this step without another hint. You should be able to do so.” Results indicated that providing feedback regarding the
SRL skill, help seeking, did in fact develop that skill, which later transferred to another similar assignment. Other studies have examined the transferability of additional SRL skills, such as self-assessment and planning/monitoring with mixed results (see Table 1.5 for a summary of those results; Roll et. al., 2014).

**Table 1.5. A summary of the findings for seven studies that measure the transferability of SRL skills taught through SRL scaffolds.** (Image copied from Roll et al., 2014, page 174).

Learning from mistakes is a desirable outcome of SRL because it is closely tied to mastery goal-orientation, positive self-reaction, and high self-efficacy (Winne & Nesbit, 2009). Therefore, systems that allow students to make mistakes, help them identify their mistakes, and then support corrective strategies lead to increased learning and better developed SRL skills, especially self-efficacy (Lorenzet, Salas, & Tannenbaum, 2005).
Other interventions designed using ITS to build self-efficacy include verbal persuasion in the form of text messages. A researcher delivered e-mail messages to students commenting on their past successes and found that self-efficacy and performance improved (Jackson, 2002). This suggests that an ITS that can easily track performance could deliver similar messages more easily. Tracking student performance and sharing that information with the learners can also help improve self-observing skills. Students with low self-efficacy have been found to struggle with accurately self-observing (Winne & Nesbit, 1998). That is to say that with metacognitively self-monitoring they underestimate the occurrence of successes. ITS can help foster self-efficacy by accurately providing feedback on the association between self-regulatory practices and outcomes (Arroyo et al., 2007). This information can be presented in the form of a graph, which explicitly shows students which strategies lead to success, allowing struggling students to become more aware of their learning by providing evidence of success.

More robust messaging interventions exist that attempt to increase motivation by developing growth mindset (Arroyo et al., 2016). Messages were delivered using a pedagogical agent, who responded to incorrect student responses while completing math questions using the ITS, MathSpring (Woolf, 2009). The content of the messages, depending on experimental condition, was correctness, growth mindset, or a combination of empathy and growth mindset.

For example, students receiving correctness messages were told after an incorrect response that their answer was wrong and offered a hint. Students receiving growth mindset messages were reminded after an error that even if they were struggling, that it is ok because they are learning something and becoming smarter. The empathy messages were delivered after students reported a negative affect and answered a question incorrectly. The pedagogical agent demonstrated empathy by validating the negative affect and then offering a growth mindset message. Results indicated that both growth mindset and empathy messages, delivered by a
pedagogical agent lead to a decrease in performance-approach goals, decrease negative affective states like boredom and frustration, and increased math learning.

Other multifaceted interventions designed to address multiple aspects of SRL have also been found to be effective (Arroyo et al., 2014). Within the ITS, Wayang Outpost (now MathSpring), different interventions were employed that specifically addressed: 1) open student models that scaffold the self-regulatory process, encouraging reflection and informed choice at key moments of boredom, 2) progress charts and tips that encourage good study habits, and 3) interventions supporting help-seeking behavior in order to improve self-monitoring and evaluation. The open-student model provided a record of student progress and achievement by topic while allowing for student choice in content. The intervention prompted students to use this feature at key moments in order to promote SRL. Progress charts serve to support the self-observation sub-processes of SRL by automatically providing formal records of progress (Arroyo et al., 2007). Lastly, help-seeking behavior was addressed by introducing a “teammate effect” where the student believed they were working with the tutor. While this did not lead to better learning, it did positively impact help-seeking behaviors in that it decreased hint abuse and overuse.

Prior research suggests that features of ITS can in fact be leveraged to provide interventions that both enhance and develop SRL skills leading to improved learning. The research conducted and presented throughout this dissertation attempts to build on self-regulated learning theory and effective use of intelligent tutoring systems to deliver interventions that develop self-regulation.

References


2 The Effect of Self-Monitoring on Learning

This chapter was originally submitted to the Artificial Intelligence in Education Conference (AIED; 2013). A shorter version was accepted and published. The complete version with additional data analysis is included here because it became the muse for the rest of my work.

Role in Self-Regulated Learning: This research addresses metacognitive self-monitoring and self-recording, which are sub-processes of SRL (see Figure 2.1). Traditionally when completing a homework assignment for math class, students might check the answers in the back of the book to monitor their performance. However, this requires additional work and is often not a standard expectation of students when doing nightly homework.

![Figure 2.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, metacognitive self-monitoring, is highlighted.](image-url)
Intelligent tutoring systems change that expectation by automatically providing self-monitoring features such as correctness feedback at the question level. This means that after students answer a question, they are immediately told if they are correct or not. This allows students to accurately monitor their performance without any additional effort. However, does this practice actually increase learning? The following randomized controlled trial was conducted to determine the impact of immediate feedback on learning when completing math homework.

The proper citation for published version of this chapter is as follows:

2.1 Background

With initiatives to provide one- to one- laptops or tablets to students in K-12 schools becoming common, it is important to identify effective uses of the technology. It has long been established that intelligent tutoring systems (ITS), with their complex feedback mechanisms can improve student learning (Anderson et al. 1995, Corbett et al. 1997, Koedinger et al. 1997). Recently, Kurt VanLehn (2011) claims that ITS can be nearly as effective as human tutors. Several studies have shown the effectiveness of ITS being used in the context of a Web-Based Homework Support (WBH) (Bonham et. al 2003, Mendicino et al. 2009, Singh et al. 2011). Similarly, VanLehn et al. (2005) have shown significant learning gains in students using the Andes Physics tutoring system in place of traditional homework. In this paper, VanLehn attributes the success of the program to the fine-grain size of the feedback, addressing the error at the step level instead of at the answer level.

However, generating content with detailed feedback and ITS features is not practical for most teachers. Computer Aided Instruction (CAI) is a reasonable alternative. The distinction is the type and granularity of feedback provided. ITS provide fine-grained, detailed and specific feedback and tutoring, often at the step or sub-step level. In contrast, CAI provides immediate feedback and hints on the answer only. Cooper
(2006) highlights how poorly conceived homework does not help learning. Therefore, if CAI effectively increases learning, then homework efficacy will be improved.

VanLehn (2011) concludes that CAI is not as effective as ITS. However, he failed to fully explore the effectiveness of WBH on learning gains when used as homework support. Kulic and Kulic (1991) also claim that CAI is not as effective as ITS. However, Kulic and Kulic was done in the 90’s before the internet was widely used in such studies. Also, Kulic and Kulic focused on the effects in the classroom and was done before the more widespread use of ITS in K-12 homework. Finally, Kulic and Kulic were interested in the immediate impact on students, not on how teachers use the data to respond to student performance.

One particular cognitive science principle we are studying is the role of feedback to students and teachers. Black and William (1998) have focused on formative assessments, with an eye on informing the teacher, and not focused on providing immediate feedback to the student. The cognitive science literature suggests that letting students practice the wrong skill on their homework even if they get feedback the next day, leads to poorer learning. Shute (2008) reviews the plethora of studies and theoretical frameworks developed around understanding the role of feedback for students as well as teachers.

Therefore, in this work we are looking at the potential of WBH to improve the efficacy of homework. WBH is one of the more unexplored areas. While its use is increasingly common at the college level, with programs such as Blackboard (www.blackboard.com) and WebAssign (www.webassign.com), it is relatively new at the K-12 level. The current study employs the use of ASSISTments, an ITS that is capable of scaffolding questions, mastery learning, and hint and bug messages. However, for this study, we ablated those features creating a “correctness only” feedback system for homework. ASSISTments is currently used by thousands of middle and high school students for nightly homework. Many teachers are typing the answers from the textbook into ASSISTments so their students can receive
immediate feedback while completing homework and the teachers can then access item reports detailing student performance.

The primary research question is can correctness-only feedback improve student learning and if so, by how much? The current research presents a randomized controlled study that examines the effectiveness of correctness-only feedback on a nightly homework assignment compared to the traditional homework method of paper and pencil without any feedback until the next day. A secondary question is, what is the effect on student learning when the teacher has access to and uses the data generated by a WBH system to review homework. Additionally, Attali and Powers (2010) recently suggested that giving students immediate feedback on a test can lead to better assessment power. If a student answered incorrectly initially, and was given a second chance they could receive a partial credit score of 2/3 instead of a zero as in the control group. We attempted to replicate these findings and expand them to determine if the partial credit score would also better allow us to predict student standardized test scores.

2.2 Methods

Participants were 65 seventh grade students in a suburban middle school in Massachusetts. They completed the activities included in the study as part of their regular math class and homework assignment. Students were assigned to conditions by blocking on prior knowledge. This was done by ranking students based on their overall performance in ASSISTments prior to the start of the study. Pairs of students were then selected and the individual students were randomly assigned to either the traditional homework (TH) (n=33) or web-based homework (WBH) (n=30) condition.

The study began with a pretest that was administered at the start of class [6]. This test consisted of five questions, each referring to a specific concept relating to negative exponents. Students were then given whole-group instruction on the current topic.
That night all students completed their homework using ASSISTments. The assignment was designed in triplets, where there were three similar questions in a row. Each triplet was morphologically similar to the questions on the pretest. To maintain ecological validity, additional and challenge questions were included throughout the homework, which resulted in a total of 20 questions. The content covered in the homework can be found in Table 2.1.

**Table 2.1. Triplet identification number with a description of the mathematical content and corresponding question numbers from the homework assignment.**

<table>
<thead>
<tr>
<th>Triplet</th>
<th>Content and Example</th>
<th>Homework Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Multiplying powers with like bases: $(3^3)(3^{-1})$</td>
<td>4, 5, 6</td>
</tr>
<tr>
<td>2</td>
<td>Multiplying powers with variable bases and coefficients: $(3x^3)(2x^{-2})$</td>
<td>8, 9, 10</td>
</tr>
<tr>
<td>3</td>
<td>Dividing powers with like bases: $(6^6)/(6^1)$</td>
<td>11, 12, 13</td>
</tr>
<tr>
<td>4</td>
<td>Dividing powers with variable bases and coefficients: $(2y^4)/(8y^1)$</td>
<td>14, 15, 16</td>
</tr>
<tr>
<td>5</td>
<td>Dividing powers with multiple variable bases: $(x^4y^2)/(x^8y^{-3})$</td>
<td>17, 18, 19</td>
</tr>
</tbody>
</table>

**Additional Questions**
Rewriting negative exponents using only positive exponents:
Rewrite $4^{-3}$ using only positive exponents.

<table>
<thead>
<tr>
<th>Homework Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 3</td>
</tr>
</tbody>
</table>

**Challenge Questions**
1.) Exponent = 0: Evaluate $12^0$.
2.) Dividing powers with multiple variable bases where an exponent is 1: $(x^4y)/(x^4y^{-2})$.

<table>
<thead>
<tr>
<th>Homework Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>7, 20</td>
</tr>
</tbody>
</table>

Students in the WBH condition were given correctness only feedback. Specifically, they were told if their answer was correct or incorrect. See Figure 2.2 to see what a student saw with a correct response and Figure 2.3 for an incorrect response.
If a student answered a question incorrectly, he was given unlimited opportunities to self-correct, or he could press the “show me the last hint” button to be given the answer. A correct response was required to proceed to the next question. Figure 2.4 shows the message students received when using the hint button.

Students in the TH condition were simply told that their answer was recorded but were not told if it was correct or not. Figure 2.5 shows the message they would receive. It is important to note that students in both conditions saw the exact same questions. The difference was the feedback received and the ability for students in the WBH condition to try multiple times before asking for a hint, revealing the answer.
2.3 Results

2.3.1 Learning Gains from Homework

Only students who completed all of the assessments and the homework were included in the study (n=63). A t-test comparing the pretest scores revealed that students were balanced at the start of the study (t(61)=0.29, p=0.78). However, an ANCOVA showed that students in the WBH condition reliably outperformed those in the TH condition on both Post-test1 (F(1,60)=4.14, p=0.046) and Post-test2.

Figure 2.5. The feedback provided by ASSISTments to students in the TH (no feedback) condition after submitting an answer.
(F(1,60)=5.92, p=0.018) when controlling for pretest score. We used CEM(2013) to calculate Hedges’ corrected effect size of .57 with a confidence interval of .07 to 1.08. See Table 2.2 for averages, standard deviations and effect sizes.

**Table 2.2. Averages, standard deviations (in parenthesis), and effect size for each dependent measure by condition. *Notes a reliable difference.**

<table>
<thead>
<tr>
<th></th>
<th>TH</th>
<th>WBH</th>
<th><em>p</em>-value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Test</td>
<td>9% (17)</td>
<td>7% (14)</td>
<td>0.78</td>
<td>NA</td>
</tr>
<tr>
<td>Post-Test1</td>
<td>58% (27)</td>
<td>69% (21)</td>
<td>0.046*</td>
<td>0.52</td>
</tr>
<tr>
<td>Post-Test2</td>
<td>68% (26)</td>
<td>81% (22)</td>
<td>0.018*</td>
<td>0.56</td>
</tr>
<tr>
<td>HW Average</td>
<td>61% (20)</td>
<td>60% (15)</td>
<td>0.95</td>
<td>NA</td>
</tr>
<tr>
<td>Partial Credit HW Score</td>
<td>61% (20)</td>
<td>81% (18)</td>
<td>0.0001*</td>
<td>1.04</td>
</tr>
<tr>
<td>Time Spent (mins)</td>
<td>22.7 (9.6)</td>
<td>23.2 (6.2)</td>
<td>0.96</td>
<td>NA</td>
</tr>
<tr>
<td>Learning Gains</td>
<td>0.03 (0.9)</td>
<td>1.73 (1.1)</td>
<td>0.0001*</td>
<td>2.21</td>
</tr>
</tbody>
</table>

A comparison of percent correct on the homework shows that students scored similarly, answering approximately the same number of questions correctly on the first attempt (F(1,60)=0.004, p=0.95). However, remember, that students in the WBH condition were given correctness feedback and then unlimited attempts to self-correct. In an attempt to replicate Attali and Powers (2010) we computed a partial credit score that gave students credit if they answered correctly without requesting a hint (Adjusted HW Average in Table 2.2). This means an adjusted homework average was computed for these students by dividing the number of questions students correctly answered on their own (without receiving the answer) by the total number of questions (20). This gives a situation similar to the partial credit model of Attali & Powers but not identical.
Attali argued their partial credit method allowed a better correlation than the more standard assessment method. We correlated students' homework percentage and the adjusted homework percent to MCAS scores (a high-stakes standardized test in Massachusetts), and found that the TH students had no correlation with MCAS, and the WBH group had high a correlation with MCAS (0.48 for traditional scoring and 0.47 for the adjusted homework score). The fact that the adjusted method did not yield a higher correlation with MCAS contradicts what Attali and Powers reported.

An ANCOVA revealed that when calculating homework performance using the adjusted homework average, students in the WBH condition performed reliably better than those in the TH condition (F(1,60)=17.58, p<0.0001). This suggests that with unlimited attempts, students are able to self-correct, allowing them to outperform their counterparts.

Furthermore, the homework assignment was constructed using triplets. Learning gains within the triplets were computed by adding the points earned on the third question in each triplet and subtracting the sum of the points earned on the first question in each triplet. Students with correctness feedback and unlimited attempts to self-correct learned reliably more while doing their homework (F(1,60)=45.72, p<0.0001).

A review of the item report further describes this difference in learning gains. As expected, students in the TH condition continue to repeat the same mistake each time the question was encountered, resulting in three wrong responses in a row. Conversely, students in the WBH condition may have repeated the mistake once or twice, but rarely three times in a row, accounting for the learning. This trend is notable in 4 out of 5 of the triplets. See Table 2.3 for the number of students answering all of the questions in a triplet incorrectly by condition.

The first thing that we want to point out is that students in the WBH condition had a significantly lower percentage correct on the first item. This is almost certainly due to the fact that students who were in that condition knew they had multiple attempts to correctly answer the question so there was no penalty for answering incorrectly on
the first attempt. Additionally, students had the hint button available, which would provide the answer at the student’s request. Eight students in the WBH condition used the hint button on the first question in triplet 1. Presumably students in the WBH condition would use the hint button when they were not sure what the answer. However, in the TH condition, there was no such button, students were more likely to take other steps when they were confused. These steps could include looking at class notes, asking a parent or calling a friend for help.

Table 2.3. The number of students answering all three questions in a triplet incorrectly by triplet identification number and condition

<table>
<thead>
<tr>
<th>Triplet Number</th>
<th>Traditional Homework Number of Students</th>
<th>Web-Based Homework Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>13</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>1</td>
</tr>
</tbody>
</table>

When looking at students in the WBH condition who could demonstrate learning (they got the first one wrong) a stunning 18 out of 22 students (80% of students) demonstrated learning. In one sense this learning benefit might be overestimated, as there were some interesting differences in response behavior between the conditions. Specifically, response time for the initial action shows that perhaps students’ approach the problems differently. We analyzed the time it took students to type in their first response and found that students in the TH condition took longer on average (121 seconds) than students in the WBH condition (89 seconds). This difference was not statistically significant. The TH condition had 34% of students take over two minutes to generate their first response while the WBH condition only had 17% of students take that long. This indicates that students in the WBH condition may have a higher percentage of incorrect first responses due to less thorough processing. We hypothesize that students in the WBH condition knew that they had
an opportunity for feedback and multiple attempts so they may have been more likely to input their first thought with or without much confidence. This would account for the higher number of students who seemingly already knew the material in the TH condition.

While the TH condition took longer to generate a first response, if you took into account the total time spent on the problem including letting the student make multiple attempts as well as asking for a hint, amazingly, the students in the WBH group took approximately the same amount of time. The TH group took on average 2 minutes and 7 seconds while the WBH group took on average 2 minutes and 27 seconds per question. (This difference was not reliably different.)

It seems that students in the TH group spend more time thinking about the problem, but the WBH group can get the problem wrong, and then use their time to learn the content. We were not expecting that correctness-only feedback was going to be time efficient. But in fact, students in both conditions spent the same amount of time to complete their homework (F(1,60)=0.002, p=0.96).

It has been argued that educational technology exacerbates the achievement gap between high and low knowledge students (Steenbergen-Hu & Cooper, 2013). To investigate this further, participants were split into “high knowledge” and “low knowledge”. We intended to use the median to split the students into two equal groups, however 46 out of the 63 students had a score of zero, therefore we split them into students with a score of zero (“low knowledge”) and students with a score greater than zero (“high knowledge”). Table 2.10 shows the means and standard deviations for learning by condition and knowledge.

A univariate analysis of variance revealed that the interaction between condition and knowledge is marginally significant (F(1, 62) = 3.83, p = 0.55). Analyzing only the “high knowledge” students (TH: N = 9, WBH: N = 8) reveals that condition was not a significant factor (F(1, 16) = 0.12, p = 0.7). However, for “low knowledge” students (TH: N= 24, WBH: N = 22), condition was a significant factor (F(1, 45) = 10.06, p =0.03). Hedges’ corrected effect size further indicates that this is a large effect (g =
0.94). This suggests that the effects of correctness feedback lead to more learning for “low knowledge” students than “high knowledge” students, perhaps closing rather than widening the achievement gap.

Table 2.4. Mean (and standard deviations) learning gains by condition and knowledge level.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean (SD)</th>
<th>N</th>
<th>Knowledge Level</th>
<th>Mean (SD)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>TH</td>
<td>58% (25)</td>
<td>33</td>
<td>High</td>
<td>47% (22)</td>
<td>17</td>
</tr>
<tr>
<td>WBH</td>
<td>74% (29)</td>
<td>30</td>
<td>Low</td>
<td>73% (27)</td>
<td>46</td>
</tr>
<tr>
<td>Total</td>
<td>66% (28)</td>
<td>63</td>
<td>Total</td>
<td>66% (28)</td>
<td>63</td>
</tr>
</tbody>
</table>

2.3.2 Learning Gains from Homework Review

To address the second research question of the effectiveness of using the data to support homework review, a paired t-test revealed that students did reliably better on Post-test2 than on Post-test1 \( (t(62)=3.87, p<0.0001) \). However, an ANCOVA revealed that when accounting for Post-test1 scores, there is not a reliable difference by condition in the gains from Post-test1 to Post-test2 \( (F(1,60)=2.18, p=0.15) \). This suggests that while reviewing the homework leads to substantially improved learning, the method of reviewing the homework does not significantly impact the amount of continued learning.

2.3.3 Observational Results

In addition to examining the effects of immediate feedback on learning, this study explored the potential changes to the homework review process the following day in class. In the traditional format of homework review, time must be spent first on checking answers and then the teacher answers student’s questions. However, we hypothesized that when teachers have access to the item report they are able to identify common misconceptions and address those ensuring that the time spent reviewing homework is meaningful.
Remember, that when reviewing the homework, students were separated by condition. The teacher recorded herself as she reviewed the homework with each group. In the following section we attempt to characterize what happened in the video segments. As usual, the teacher reviewed the item report in the morning to determine which questions needed to be reviewed in class. The item report shows that triplet 1 showed a common misconception when multiplying powers with like bases. While the item report shows that students learned from the feedback, the teacher still felt it was important to highlight and discuss the error in multiplying the bases of the powers together. Therefore the teacher highlighted question 4.

This was especially important because in triplet 2, students incorrectly applied this concept. Specifically, you can see in the item report that only 39% of students initially got this type of question right (multiplying powers with coefficients and variables). However, learning took place as 68% got the next similar question right. It was therefore puzzling to see that on the third question in that triplet (question number 10), only 45% got the question right. Upon investigating the question, the teacher was able to identify the misconception and therefore addressed it with the class. Students learned in the prior triplet not to multiply the bases together. However, in this problem (5a^3)(5a^5) students didn’t realize that they could multiply the coefficients, 5 and 5 together. You can see in the video that the teacher highlights the difference between these types of problems.

The third triplet showed adequate learning and therefore the teacher chose not to review any of these questions. However, the 4th triplet proved to be the most challenging and showed little learning. Therefore, the teacher chose to review the first question of the triplet (question number 14.) The teacher was able to identify the common mistakes, which were improperly subtracting the negative exponents as well as dividing the base. Because the next question had the poorest performance on the assignment, the teacher also chose to review question number 15 and highlight the importance of subtracting negative exponents carefully.
The final triplet also showed adequate learning and therefore the teacher decided not to review any of those questions. Additionally, questions 1, 2, and 3 were introductory questions and performance was above 90% on each question so the teacher did not address these. Similarly, questions 7 and 20 were not part of any triplet and were challenge questions. Therefore the teacher did not take class time to explain those.

We designed the experiment with ecological validity in mind. That is to say, we wanted the teacher to naturally review the homework, giving students enough time to ask questions. The hope was that approximately the same amount of time would be spent in each class and by each condition. We were disappointed to find that the classes and conditions varied greatly in the amount of time spent going over the homework. As you can see in Table 2.5, half of the sections took over nine minutes to review the homework while two of the sections in the TH condition and one in the WBH condition spent substantially less time. This is a threat to the validity of drawing statistical inferences, but given the desire to maintain realistic homework review conditions, these inconsistencies highlight important differences in the homework review methods. We describe these differences in the following sections.

Table 2.5. Class time spent (in minutes and seconds) on homework review and the specific homework question numbers covered for each class by condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Time Spent</th>
<th>Question Numbers Reviewed</th>
<th>Triplets Addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Homework</td>
<td>6 mins 48 secs</td>
<td>15, 17, 15</td>
<td>4, 5, 4</td>
</tr>
<tr>
<td>Traditional Homework</td>
<td>9 mins 26 secs</td>
<td>7, 11, 10, 15, 14, 16, 8</td>
<td>3, 2, 4, 4, 4, 2</td>
</tr>
<tr>
<td>Traditional Homework</td>
<td>5 mins 41 secs</td>
<td>8, 4</td>
<td>2, 1</td>
</tr>
<tr>
<td>Web-Based Homework</td>
<td>9 mins 29 secs</td>
<td>4, 10, 14, 15, 16</td>
<td>1, 2, 4, 4, 4</td>
</tr>
<tr>
<td>Web-Based Homework</td>
<td>9 mins 26 secs</td>
<td>4, 10, 14, 15, 12</td>
<td>1, 2, 4, 4, 3</td>
</tr>
<tr>
<td>Web-Based Homework</td>
<td>5 mins 51 secs</td>
<td>4, 10, 14, 15</td>
<td>1, 2, 4, 4</td>
</tr>
</tbody>
</table>
An observational analysis of the video recordings of the teacher reviewing the homework revealed that while the time spent in the WBH condition was often longer than the TH condition, it was also far more focused than in the TH. Specifically, when students were in the TH condition, on average 1 minute passed before any meaningful discussion took place. Whereas, when students were in the WBH condition, homework review began immediately with the teacher reviewing the 4 most troublesome problems.

Other notable differences in the type of review include the number of questions answered. In the TH condition, 2 classes saw 3 questions each and one saw 7. However, in the WBH condition each class saw 4 targeted questions and 2 classes requested 1 additional question. The variation in question types also is important to note. The teacher was able to ensure that a variety of question types and mistakes were addressed whereas in the TH condition students tended to ask the same types of questions or even the same exact question that was already reviewed. Additionally, students in the TH condition also asked more general questions like “I think I may have gotten some of the multiplying ones wrong.” You can see in Table 2.5 that in one TH condition only multiplication questions were addressed when clearly division was also a weakness and similarly, another TH condition only asked questions about division. This accounts for much of the variability in overall review time.

In listening to the comments made by students, it appears that the discussion in the TH condition was not as structured as the WBH. Not all students had their work and therefore couldn’t participate in the review. One student said, “I forgot to write it down.” Another said, “I left my work at home.” Because students were asking questions directly to the teacher, who was answering that specific student, we suspect that only the student who asked the question was truly engaged. In fact, one student said, “I was still checking and couldn’t hear” which led to the teacher reviewing a question twice. In the WBH condition, the teacher used the information in the report, such as percent correct and common wrong answers to engage the entire
class in a discussion around misconceptions and the essential concepts from the previous question.

Another notable difference is the completeness of the review. In the TH condition, the review was dominated by student directed questions. This means that each class experienced a different review and the quality of that review was directly dependent on the engagement of the students. Conversely, in the WBH condition, all 3 classes were presented with the same 4 troublesome questions and common mistakes. Additional questions were also reviewed when asked.

2.3.4 Other general observations or comments

In the TH condition the teacher asked students if they had questions on an average of six times, repeating “Which questions gave you trouble?” “Which questions did you get wrong?”.

One student asked about question 7 but called it question 8. In the TH condition the teacher said, “Oh this is the one we were just looking at.” In the WBH condition a student asked for the class average showing a deeper understanding of data collection and how that can be used in the learning process. In the WBH condition some of the discussion surrounded identifying the mistakes students made to arrive at the incorrect response.

2.3.5 Student Survey Results

Following participation in this study, students were questioned about their opinions. We want to acknowledge that students might have been telling the teacher what she wanted to hear: the whole classroom of students had been using ASSISTments for months and the teacher had told them on multiple occasions why it was good for them to get immediate feedback. So with that caveat, we share the following results. 86% of students claim “to prefer using ASSISTments to a worksheet when completing their homework
In response to the question “Do you prefer to do your homework on ASSISTments or a worksheet?”, 84% said they learn more when they use ASSISTments to complete their homework. 66% mistakenly think that it takes longer to complete their homework when using ASSISTments and 44% feel that they get frustrated when using ASSISTments to complete their homework. However 73% say that their time is better spent using ASSISTments for their homework than a worksheet. Therefore, it is important to remember that while the learning benefits are profound and students prefer a web-based system, there is a sense of frustration that must still be addressed. Maybe we can change the software to add more AIED feature, or maybe learning requires some levels of frustration.

2.4 Conclusions

This papers’ contribution to the literature is helping the AIED field think about reasonable effect size estimates for comparing AIED interventions to business and usual control conditions. VanLehn relied on a subset of the papers that Kulic and Kulic reviewed to estimate the effect of “answer-based tutors” to be 0.3 std. Those studies compared computer to classrooms instruction. But a different, and arguably more relevant, comparison is comparing traditional homework to computer supported WBH conditions that also allows for going over the homework using the data. Both elements seem to have some utility.

This randomized controlled study suggests that simple correctness-only feedback for homework substantially improves learning from homework. Having the data to do a more effective homework review benefited the teachers in the expected direction (but was not reliable). But taken together (immediate feedback at night, and a arguable smarter homework review driven by the data) the effect size of 0.56 seems much larger than the 0.3 standard deviations that Vanlehn used for answer based tutoring. Of course, the large 95% confidence interval of [.07 to 1.08] tells us we need more studies. But if this .56 effect size is true, it is almost double the Kulic and Kulic estimate of 1990’s computer based approaches. It .56 is a reliable effect size it would be raising the bar for any AIED interventions that include immediate feedback and reports to teachers.
Personally, we are excited to raise the bar by adding some AEID-like features like 1) mastery learning on problems the students are having difficulty with, 2) cognitively diagnostics “bug messages for common wrong answers, and 3) step by step hints and other features of intelligent tutoring systems. We encourage the field to use WBH as a useful control condition. We suspect that the effect size of such a system could be very high (VanLehn has reported effect sizes of Andes compared to traditional homework to be over 1 standard deviation.

Furthermore this paper attempts to replicate Attali and Power’s reporting of an assessment power benefit but reports that this intuitive idea did not work out as their past work suggested. This serves as a useful data point for that group of educational measurement professionals to ponder.

Caveats: the participants in the current study were all advanced middle school students. Therefore it would be necessary to replicate this study across a broader range of student abilities to determine if these effects are generalizable. Additionally, the correctness feedback is confounded with the unlimited attempts provided on the homework assignment. Therefore, it would be interesting to see if it’s simply the correctness feedback that contributes to learning or if the impact stems from the unlimited attempts to self-correct. Finally, to address the secondary research question of the effectiveness of using that data and item report to enhance homework review, a more complicated research design would be required. Specifically, in the present study, the effect of the homework review was confounded with already improved learning that resulted from having correctness feedback. A two-by-two design where we vary both immediate feedback and the factor of going over the homework with the data should be considered in future research. In this fast-paced educational world, it is important to ensure that time spent in class and on homework is as beneficial as possible. This study provides some strong evidence that web-based homework systems that provides correctness-only feedback are useful tools to improve learning without additional time.
The utilization and purpose of homework in today’s schools are debatable (Foyle & Bailey, 1986). If homework is simply meant to practice what students already know, then feedback might not be necessary. However, how do students know that they are completing their homework correctly? What if a student completes his entire assignment wrong and has practiced the same mistake many times? If the homework assignment covers a new concept that has not yet been mastered or is an extension of the current topic, then student mistakes are expected. This randomized-controlled study clearly shows the benefits of correctness-only feedback while completing homework therefore lending support for its use for nightly homework. It is no surprise that students who were given correctness feedback while they completed their homework outperformed those who received no feedback at all. In the traditional paper and pencil model that relies on teacher feedback the next day (or in some cases several days later), it is assumed that the feedback provided and any homework review will provide sufficient intervention to maximize learning. However, this study reveals that this is not the case. Students who received correctness feedback and targeted teacher review of the homework far outperformed their previously mentioned counterparts. That is to say that teacher review of the homework is unable to recreate the learning that results from feedback while students initially do their homework.

In fact, the results indicate that correctness feedback while students complete homework is as effective as receiving no feedback and teacher review of the homework. This suggests that to save time, teachers may not even need to review the homework if students have access to web-based homework systems.

Additionally, it is interesting to note that receiving feedback and making multiple attempts, over 100 in some cases, did not take any longer than completing the homework without feedback. That is to say that for the same time spent learning, students using a WBHS learned 13% more than students using the traditional homework model. It is for all of these results that this study suggests that all students should be using web-based-homework systems with correctness feedback for nightly homework.
References


Kelly, K (2012f) Survey Results stored here http://web.cs.wpi.edu/~nth/PublicScienceArchive/KellySurvey.csv


CEM (2013) Downloaded from here http://www.cemcentre.org/evidencebased education/effectsizecalculator


3 Motivating the Unmotivated

This chapter was published as part of the 2013 Florida Artificial Intelligence Research Society (FLAIRS) conference. It is included here because the intervention presented in this randomized controlled trial became the basis for later interventions.

**Role in Self-Regulated Learning:** This randomized controlled trial was an attempt to increase motivation in some of my students who I perceived to be lacking motivation. I realized that the learning gains noted in Chapter 2 were not observed consistently in all of my 7th grade math classes. Self-regulation theory recognizes that goal orientation and task value/interest impact learning (see Figure 3.1).

![The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, self-motivational beliefs, specifically outcome expectations, and task value/interest, are highlighted.](image-url)

**Figure 3.1.** The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, self-motivational beliefs, specifically outcome expectations, and task value/interest, are highlighted.
Therefore it was hypothesized that students with performance-based orientations and low task value/interest would not take advantage of the correctness feedback in order to reap the learning benefits observed in my advanced classes. The intervention included in this study was designed to determine if it is possible to modify students’ beliefs. If so, do these shifts in beliefs have an impact on motivation and ultimately increase learning?

The proper citation for this chapter is as follows: 

3.1 Introduction

In recent years, connections between complex learning and emotions have received increasing attention in the intelligent tutoring literature (Arroyo et al., 2009; Conati & Maclaren, 2009; Forbes-Riley & Litman, 2009; Kort, Reilly, & Picard, 2001; Robison, McQuiggan, & Lester, 2009; Woolf et al., 2010). Deeper understanding of affect-learning connections is needed to design more engaging educational technologies, creating the potential for affect-sensitive intelligent tutoring systems and more engaging educational games (De Vicente & Pain, 2002; Graesser, Jeon, & Dufty, 2008; Litman & Silliman, 2004; Sabourin, Mott, & Lester, 2011). The focus on affect is motivated by the realization that affect and cognition are complementary processes in learning environments.

In light of the affect-cognition relationship, a number of research groups have been addressing the problem of building learning environments that detect and respond to affective states such as boredom, confusion, frustration, and anxiety (Afzal & Robinson, 2009; Burleson & Picard, 2007; Chaffar, Derbali, & Frasson, 2009; Conati & Maclaren, 2009; D’Mello & Graesser, 2010; D’Mello et al., 2010; Forbes-Riley, Rotaru, & Litman, 2008; Robison, McQuiggan, & Lester, 2009; Woolf et al., 2010). These systems use state-of-the-art sensing technologies and machine learning methods to automatically detect student affect by monitoring facial features, speech
contours, body language, interaction logs, language, and peripheral physiology (e.g., electromyography, galvanic skin response) (see (Calvo & D'Mello, 2010) for an overview). These affect-sensitive systems then alter their pedagogical and motivational strategies in a manner that is dynamically responsive to the sensed affective states. Some of the implemented responses to student affect include affect mirroring (Burleson & Picard, 2007), empathetic responses (Woolf et al., 2010), and a combination of politeness, empathy, and encouragement (D'Mello et al., 2010).

The above studies represent some of the contemporary approaches to recognizing and responding to affect. Some are multi-million dollar efforts. This study takes a different approach in attempting to improve student affect using teacher-generated video messages. In addition to being considerably more cost-effective to produce, such videos might have additional benefits of personalization because the message is delivered by the student’s teacher rather than an unfamiliar animated agent (Cordova & Lepper, 1996). Our study does not attempt to create a sophisticated emotional model of the student. Instead we attempt a novel approach of leveraging the “skill builder” infrastructure in the intelligent tutoring system, ASSISTments to deliver the motivational interventions.

The present interventions are grounded in the control-value theory of academic emotions (Pekrun et al., 2006) and on perspectives that highlight the importance of students’ attitudes on the malleability of intelligence (Dweck, 2006).

According to Pekrun’s control-value theory of academic emotions, subjective appraisals of control and value of a learning activity give rise to student affect (Pekrun, 2010; Pekrun et al., 2006; Schutz & Pekrun, 2007). Subjective control pertains to the perceived influence that a learner has over the activity, while subjective value represents the perceived value of the outcomes of the activity. Engagement is increased when task value is high, but boredom dominates when learners feel like they are being forced to exert effort on an activity they do not value (Hulleman et al., 2008; Pekrun, 2006; Pekrun et al., 2010). Therefore, one would expect that students completing a homework assignment might experience boredom
when task value is perceived as being low or when problem difficulty is considerably below or above their skill level.

In a related vein, Dweck and colleagues discovered that people with trait-based beliefs in their own competencies (also termed the “entity theory of intelligence” or the “fixed mindset”) are much more likely to disengage when they are confronted with challenges and threats of failure (Dweck, 2002; 2006). In contrast, individuals who believe that their competency can be developed with practice and effort (i.e., the “incremental theory of intelligence” or the “growth mindset”) are much resilient to failure and its resultant negative effect (such as frustration or anxiety). Past research suggests that messages that help students change their attributions of failure from external, fixed, and uncontrollable factors to internal, malleable, and controllable factors are effective at re-engaging students (Shores & Smith, 2010;). Students are expected to become frustrated and disengage when they perceive that the learning task is out of their control due to adherence to entity theories of intelligence.

In line with these theories, one proposed intervention consists of displaying a video message designed to help learners reappraise the learning task, so as to increase its perceived value (value-video condition). The message reminds learners that although they may see little value in the specific current task (e.g., surface area of 3-dimensional figures), each task helps to broaden and build upon their math skills, get a good grade and progress to the next grade level, as well as has the potential to be used in a future career or to improve general problem solving abilities. The other intervention incorporates a motivational video message to the students in order to help them alter their attributions of failure, and promote a growth mindset. This intervention emphasizes the fact that failure is part of learning and if the student continues to work hard, they will be met with success (control-video condition). A third condition, called the no-video condition serves as a control.

A second related theoretical grounding of this work is related to the personalization of learning. Cordova and Lepper’s (1996) classic study showed that by personalizing an intelligent tutor with information like the student’s favorite food and the names of
their classmates, motivation and learning increased. Can we leverage this personalization effect by having the students’ teachers record these different motivational messages instead of having them be delivered by an animated pedagogical agent or as plain text?

These interventions were tested in the context of a real math class and over the course of an actual homework assignment with ASSISTments, a math ITS (Mendicino, Razzaq & Heffernan 2009), thus providing ecological validity. ASSISTments allows students to complete problem sets for homework, providing tutoring and hints at the student’s request. The pre- and post-tests were administered during class before and after the targeted homework assignment with ASSISTments, respectively. These students were accustomed to using ASSISTments, so, while the embedded video was novel, the homework task was not.

The primary research question was, can a teacher-generated video help change students' perceptions of the value of math or encourage a “growth mind-set”? If so, what effect does this change have on perseverance as measured by homework completion?

3.2 Study 1

This study attempted to use an intelligent tutoring system to deliver a motivational message in the form of videos starring the teacher. The videos targeted either students’ value in or their control over the learning activity. The effects of the videos on students’ perceived value, control, and learning gains were measured and analyzed.

3.2.1 Experimental Design

Twenty-four 7th grade students, in a suburban school in Massachusetts, participated in this study as part of their homework in their math class. In an attempt to balance the groups, students were blocked into groups based on prior knowledge. This was
done using students’ current averages in math class, ordering them from least to
greatest, taking three consecutive students at a time and randomly assigning them to
the no-video condition (n=8), control-video condition (n=8), or value-video condition
(n=8).

All students were given a pre-test and a pre-conception survey in class. The pre-test
consisted of five content questions (adding and subtracting fractions). The actual pre-
test can be found in Kelly (2012a). The pre-conceptions survey consisted of five
questions to measure attitudes towards value and control. The survey questions
were adapted from the Academic Emotions Questionnaire (Pekrun, Goetz & Perry
2005) and used a 5-point Likert scale. The specific statements were:

1. With hard work, most anyone can be good at math. (control)
2. If I apply myself, I can improve my grade in math. (control)
3. I like learning math skills. (liking)
4. Math skills taught in school are valuable. (value)
5. I think that knowing the math skills taught in school will be good for me
when I am older. (value)

Students were then given a homework assignment in ASSISTments. As part of the
students’ typical routine, this assignment was completed at home as their math
homework. The homework started with three content related questions that were
morphologically similar to the pre-test. Students in the experimental condition, who
got at least one of these questions wrong were given a link to one of two possible
videos on YouTube. The videos were of their math teacher delivering a motivational
intervention. The control-video had the teacher explaining that if the student works
hard, they can be successful. The value-video had the teacher explaining how
important math homework is to learning as well as to their future. (See Kelly (2012b)
for the actual videos). The assignment continued with content related questions until
the student independently and correctly answered three consecutive questions.
Correctness feedback was provided and hints were available upon student request to
assist in answering a question. Finally, after completing the content questions,
students were given the same five survey statements that had been asked during the pre-test. To see the entire assignment, as a student would experience it, see Kelly (2012c).

The next day, students were given a post-test in class. The post-test consisted of five morphologically similar, content related questions. Five days later, a retention post-test was given consisting of five content related questions (see Kelly (2012d) for the items). Students also completed the 5-item survey at both these testing points.

### 3.2.2 Results

Students who answered the first three questions correctly on the homework assignment did not receive the intervention and were therefore not included in the analysis (n=5). Only students who completed all of the tasks were included (n=16).

The results of the survey questions are shown in Table 3.2. An ANOVA showed a significant main effect of value-video ($F(2,13)=4.51, p=0.03$) after completing the homework assignment. Specifically, post hoc tests showed that students viewing the value video showed marginally higher student appraisals of the perceived value of math than the students in the no video condition immediately after the homework assignment ($F(2,13)=3.01, p=0.084$). This effect was stronger the following day on the immediate post-test ($F(2,13)=4.30, p=0.037$) and, to a lesser degree, several days later on the delayed post-test ($F(2,13)=2.09, p=0.12$).

There were no significant differences of meaningful trends for the control-video, which suggests that this intervention was less successful than the value-video.

Due to the small sample size and scale of the intervention, differences in learning gains were not expected. However, for completeness, mean scores were calculated and are presented in Table 3.2. As expected, averages were not reliably different on the immediate post-test ($F(2,13)=0.036, p=0.9$), or the delayed post-test ($F(2,13)=0.067, p=0.9$).
Table 3.1. Means and standard deviations (in parenthesis) for survey questions on each task by condition

<table>
<thead>
<tr>
<th></th>
<th>Value Questions</th>
<th>Control Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-test</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Video</td>
<td>3.6 (0.6)</td>
<td>4.2 (0.4)</td>
</tr>
<tr>
<td>Value Video</td>
<td>3.8 (0.8)</td>
<td>4.4 (0.4)</td>
</tr>
<tr>
<td>Control Video</td>
<td>4.3 (0.3)</td>
<td>4.4 (0.5)</td>
</tr>
<tr>
<td><strong>Homework</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Video</td>
<td>3.6 (0.6)</td>
<td>4.3 (0.4)</td>
</tr>
<tr>
<td>Value Video</td>
<td>4.4 (0.7)</td>
<td>4.4 (0.2)</td>
</tr>
<tr>
<td>Control Video</td>
<td>4.1 (0.5)</td>
<td>4.4 (0.5)</td>
</tr>
<tr>
<td><strong>Immediate Post</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Video</td>
<td>3.3 (0.6)</td>
<td>4.1 (0.5)</td>
</tr>
<tr>
<td>Value Video</td>
<td>4.3 (0.7)</td>
<td>4.2 (0.5)</td>
</tr>
<tr>
<td>Control Video</td>
<td>3.5 (0.9)</td>
<td>3.8 (0.9)</td>
</tr>
<tr>
<td><strong>Delayed Post</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Video</td>
<td>3.4 (0.5)</td>
<td>4.2 (0.5)</td>
</tr>
<tr>
<td>Value Video</td>
<td>4.1 (1.0)</td>
<td>3.8 (1.1)</td>
</tr>
<tr>
<td>Control Video</td>
<td>3.1 (0.9)</td>
<td>3.8 (0.9)</td>
</tr>
</tbody>
</table>

Table 3.2. Mean and standard deviation (in parenthesis) for percent of correct answers on knowledge tests

<table>
<thead>
<tr>
<th>Test Performance</th>
<th>Control Video</th>
<th>Value Video</th>
<th>No Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>65% (21)</td>
<td>60% (35)</td>
<td>73% (34)</td>
</tr>
<tr>
<td>Immediate Post</td>
<td>73% (18)</td>
<td>75% (27)</td>
<td>93% (9)</td>
</tr>
<tr>
<td>Delayed Post</td>
<td>65% (30)</td>
<td>63% (23)</td>
<td>85% (17)</td>
</tr>
</tbody>
</table>
3.3 Study 2

Based on the results of Study 1, we replicated the study with a larger sample to examine potential learning gains as a result of these reappraisals.

3.3.1 Experimental Design

Seventy-six 8th grade students, in a suburban school in Massachusetts, participated in this study as part of their homework in their algebra math class. As in study 1, groups were balanced using blocking on prior knowledge, and were then randomly assigned to a no-video condition (n=29), control-video condition (n=38), or value-video condition (n=38).

Students were given a homework assignment in ASSISTments. The first six questions of the assignment were used as a pre-test to assess how fluent students were in finding the surface area of cylinders. Students who answered all six questions correctly (n= 13) were then given survey questions but were not given the intervention and were therefore not included in the study.

Following the pre-test, students in the experimental condition were given a link to one of two possible videos on YouTube. The videos were the same from Study 1. Students were then asked to indicate how many questions they think they answered correctly. Students in the no-video condition were simply asked this question and were not given a link to a video.

Students then began the homework assignment. There were three sections in the assignment. The first required students to find the surface area of a cylinder given the diameter or radius, and height or lateral surface area. Students completed this section by answering three questions in a row correctly. The second section required students to, given the surface area, find the height, radius, diameter base area, or lateral surface area. Again, students completed this section by answering three questions in a row correctly. ASSISTments provided correctness feedback as well as
offered hints on demand to students. The third section of the assignment consisted of eight survey questions to measure students’ attitudes towards value and control. A 5-point Likert scale was used. Based on student feedback from Study 1, we limited the number of times we asked students these questions and therefore included additional questions for a more robust measure (see Kelly 2012e). The specific questions were:

1. With hard work most anyone can be good at math. (control)
2. If I apply myself I can be good at math. (control)
3. I like learning about math. (liking)
4. Math skills taught in school are valuable. (value)
5. I think that knowing the math skills taught in school will be good for me when I am older. (value)
6. If I don’t understand the course material, it is because I didn’t try hard enough. (control)
7. Some people just cannot be good at math no matter what they do. (control)
8. I do not enjoy learning about math. (liking)

The following day in class, students were given a post-test. This test consisted of six questions that were morphologically similar to the ones given on the pre-test. Finally, six days later students were given a retention post-test, again consisting of six morphologically similar questions. Following the retention post-test, students were given the same eight survey questions as during the homework assignment. To preview an actual post-test see Kelly (2012f).

3.3.2 Results

To analyze homework completion rates, only those students who attempted the homework assignment were included in the analysis (n=60). This does not include the students who answered the first six questions correctly (n=16) as they never received the intervention. Students in the value-video condition had substantially higher homework completion rates (95%) when compared to students in the control-
video (70%) and no-video conditions (69%). A Chi-Square test revealed that these differences were marginally significant ($X^2(2)=4.611, p=0.099$).

For the analysis regarding survey questions and learning gains, only students who completed all three tasks, and who received the intervention (in the case of the experimental condition) were included in the analysis ($n=27$).

Responses to the control questions were averaged together, with question 7 being reversed scored, and the value questions were averaged together for the analysis. There was no condition effect for the value-video, thereby failing to replicate a finding from Study 1. It is important to note, that students had to complete the homework to receive the survey questions. Therefore, there is the potential for selection bias in these results; specifically, if students that got the value-video were more likely to complete their homework, the sample of students in the value condition might have a different perception of the survey questions. This possible source of bias might account for our failure to replicate the findings from Study 1 pertaining to the value-video.

**Table 3.3. Means and standard deviations (in parenthesis) of survey questions on each task by condition and for percent of correct answers on knowledge tests**

<table>
<thead>
<tr>
<th>Survey Questions</th>
<th>Control Video</th>
<th>Value Video</th>
<th>No Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>HW Control</td>
<td>4.0 (0.5)</td>
<td>3.6 (0.3)</td>
<td>4.1 (0.5)</td>
</tr>
<tr>
<td>HW Value</td>
<td>3.7 (1.1)</td>
<td>4.1 (0.6)</td>
<td>4.4 (0.8)</td>
</tr>
<tr>
<td>Post-Test Control</td>
<td>3.6 (0.8)</td>
<td>3.4 (0.5)</td>
<td>3.8 (0.3)</td>
</tr>
<tr>
<td>Post-Test Value</td>
<td>3.9 (0.9)</td>
<td>3.9 (0.8)</td>
<td>4.1 (0.7)</td>
</tr>
</tbody>
</table>

**Test Performance**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>65% (22)</td>
<td>64% (21)</td>
<td>56% (33)</td>
</tr>
<tr>
<td>Immediate Post</td>
<td>87% (12)</td>
<td>77% (23)</td>
<td>92% (9)</td>
</tr>
<tr>
<td>Delayed Post</td>
<td>95% (5)</td>
<td>89% (13)</td>
<td>95% (4)</td>
</tr>
</tbody>
</table>
Similar to Study 1, no significant differences were found in percent correct on the immediate post-test \( (F(2,25)=1.67, \ p=0.20) \) or the delayed post-test \( (F(2,25)=1.27, \ p=0.30) \). See Table 3.3 for average perceptions and test scores by condition.

3.4 Contributions and Discussion

This study suggests a novel use of ITS in a highly ecologically valid situation (i.e., use in a real math classroom for homework). Specifically, providing value-focused videos from their own teacher might increase students' motivation and problem completion rates. We acknowledge that the empirical contributions of this work are modest. Instead, the real contribution of this paper is to develop a proof-of-concept of the possibility to implement cost-effective motivational interventions in intelligent tutoring system. The long-term goal of this research is to identify interventions that work (from both a cognitive and motivational standpoint) and to deploy these at the appropriate time. For example, we envision a system that identifies when students are struggling and displays different types of pre-recorded messages. There will need to be different messages for different anticipated problems; some might be content related but others might be focused on motivation. We are in the process of augmenting ASSISTments with the infrastructure to support teachers posting their own messages for common situations, and these messages can be accompanied with videos.

3.4.1 Discussion and Future Research

The results of the two studies in this paper were suggestive, and by no means conclusive. We anticipated the difficulty in obtaining statistically significant effects given that our intervention consisted of a 60-second video and the sample sizes were quite small. Clearly, future studies should be conducted implementing more substantial and long-term interventions.
Additionally, matching the interventions to the specific needs of the student might yield more impressive outcomes. Specifically, a learner who is struggling and therefore disengaged due to low appraisal of control (despite valuing mathematics) might receive supportive messages or a video to help the student feel a greater sense of control. This intervention might succeed in alleviating boredom in the short-term, but boredom might emerge as the session progresses (value decreases) and as the topic is mastered (control increases). In this situation, additional messages, either text or video, may be required. More research is needed by the field to see if this is an effective way of improving intelligent tutoring systems by personalizing motivational feedback to individual students goals and needs.

It is also important to compare the effectiveness of teacher-generated messages versus other motivational messages. It could be that teacher created videos are more effective due to the fact they come from the students' teacher rather than some generic source (maybe a paid professional actor) or an animated pedagogical agent. An added benefit might be an increase in teachers' willingness to adopt the ITS if they feel they can override system default messages with their own personalized messages.

In conclusion, Study 1 showed promising results in the ability to alter student perceptions of the value of math by showing a short, teacher-generated video during a homework assignment. Study 2 showed that this intervention might also increase motivation and therefore homework completion rates, which, over time could lead to improved learning. These studies were intended to serve as a proof-of-concept to test the idea that there might be potential benefits for including teacher-delivered motivational interventions within the context of mathematics homework completion with an intelligent tutoring system. Further research is needed to refine the interventions, augment the infrastructure to afford easy production of additional messages for a variety of situations by a variety of teachers, understand how to tailor messages to different situations and learners, and to study the long-term efficacy of this approach to facilitate learning by increasing motivation to learn.
References


  a. Pre-Test: http://www.webcitation.org/6CCmWVad5
  b. Control Video: http://www.webcitation.org/6CCmq7KpZ & Value Video: http://www.webcitation.org/6CCmiK5pb
  c. Homework Assignment: http://tinyurl.com/cbze4nr
  d. Post-Test: http://www.webcitation.org/6CCoZrovS

Study 2 Materials. Accessed 11/15/12
  e. Homework Assignment: http://tinyurl.com/c6k64ys
  f. Post-Test: http://www.webcitation.org/6CCoqgspN


Woolf, B., Arroyo, I., Muldner, K., Burleson, W., Cooper, D., Dolan, R. (2010). The effect of motivational learning companions on low achieving students and students with disabilities In J. Kay & V. Aleven (Eds.), *Proceedings of 10th*
4 Growing a Growth Mindset

The randomized controlled trial presented here was the basis of my dissertation proposal and was intended originally to become my full dissertation. The intervention failed to produce significant results due to the inherent challenges in measuring and developing self-regulated learning.

Role in Self-Regulated Learning: Self-regulation theories suggest that learning is indirectly linked to mindset (see Figure 4.1). One’s self-motivational beliefs impact one’s behavior during a learning activity, which will then impact learning outcomes. Based on prior research (Blackwell, Trzesniewski & Dweck, 2007; Karumbaiah et al., 2017; Kelly et al., 2013), it appears to be possible to shift student beliefs.

![Figure 4.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The present research focuses on addressing several aspects of this cycle while evaluating the impacts on others. Specifically, the intervention attempts to shift students’ self-motivation beliefs and causal attribution. The impact on help seeking behaviors and learning were measured.](image)
Therefore, the present research employs an intervention that combines several elements from prior successful work. It was designed to address multiple sub-processes of SRL including; goal-orientation, mindset, task value/interest, help-seeking, self-recording, and causal attributions in an attempt to modify student behavior during learning tasks in a math class, as well as increase overall learning.

4.1 Background

Self-Regulated Learning (SRL) requires a complex interaction between many sub-processes (Zimmerman & Campillo 2003). Effective self-regulation leads to greater success in a variety of domains (citation). Prior research has shown that developing self-regulation sub-processes specific to education, like goal-orientation, task value, and metacognitive self-monitoring, is not only possible, but also leads to improved academic performance (Blackwell, Trzesniewski & Dweck, 2007; Kelly et al., 2013). Furthermore, there is evidence that it is possible to leverage the features of Intelligent Tutoring Systems (ITS) to deliver successful interventions (Arroyo et al., 2014).

The randomized controlled trial presented in this chapter employs a unique intervention based on self-regulated learning theory and is designed to develop multiple SRL sub-processes. What makes this intervention so unique is that it fuses together several components from interventions that have already been proven successful and delivers it through the Intelligent Tutoring System, ASSISTments. The basis of the present intervention is work done by Blackwell and colleagues (2007). In that study, students receiving the intervention participated in eight tutorial sessions (see Figure 1.4 for details) where they learned about brain research, specifically growth mindset. Recognizing that the time commitment for that intervention is somewhat prohibitive, the present study will employ an abbreviated tutorial, which has also been proven to be effective (Aronson, Fried & Good, 2002). In addition to participating in interactive tutoring sessions, the present intervention also requires students to respond to open-ended questions inspired by those used in Aronson et
al. (2002). The overall structure of the present intervention is very similar to the original in that students will receive direct instruction on growth mindset and respond to a series of questions over several class sessions.

In addition to growth mindset messages, the intervention will also contain teacher-created motivational videos designed to increase students’ perceptions of task value and reinforce the growth mindset message. These videos were inspired by Kelly et al. (2013), however they are updated versions of the videos created specifically for this intervention. Students will respond to open-ended questions following each video to ensure student engagement, which Yeager & Walton (2011) argue is a critical component to an effective intervention. Specifically, they claim that interventions based in persuasion must actively engage the participant so as to ensure that they are not just passively receiving the message.

Beyond the tutorial sessions, students receiving the intervention will also receive targeted growth mindset messages while completing math practice. These messages are inspired by Arroyo et al. (2016). While the original intervention delivered these messages via text and a pedagogical agent, the present intervention will only use text messages. Additionally, the messages included in the present intervention also include causal attribution features that help students associate their success with effort and persistence through struggles and associate their failures with part of the learning process.

Based on SRL theory, the present study delivers a multifaceted intervention that is designed to develop specific SRL sub-processes (see Figure 4.2 for the design model). The target sub-processes include goal-orientation, growth-mindset, task value, self-efficacy, and causal attribution. Based on the cyclical nature of SRL theory it is hypothesized that shifts in these motivational beliefs will impact other SRL sub-processes like help-seeking behavior (Zimmerman & Campillo 2003). Specifically, students will use help features wisely as measured by the time spent after requesting assistance. Additionally, students will use the automatic self-recording feature more
effectively as measured by the time spent after making an error. Based on prior research findings, other hypothesized outcomes from shifts in self-beliefs include decrease in negative affect, specifically frustration (Arroyo et al., 2016), willingness to accept a challenge (Dweck, 2008; Arenas, Tabernero, & Briones, 2006) and increased persistence as measured by assignment completion (McClendon, Neugebauer, & King, 2017). Lastly, as a result of these shifts in behavior, we expect to see increased learning and higher performance.

Figure 4.2. Logic model for experimental design.
4.2 Experimental Design

4.2.1 Participants
There were 99 students enrolled in the 7th grade math class where this study was run. However, 13 students did not participate because they were either absent throughout most of the study or received part of their math instruction in a small group setting. Students attended a public school in an urban school district. The researcher was also their math teacher. The students were familiar with ASSISTments, the web-based program that was used to conduct the study, as they had been using it for nightly homework.

4.2.2 Materials

The following materials were used in this study. See Kelly (2018) to view them as a student.

Self-Motivation Beliefs Survey: To measure the constructs of goal orientation, mindset and task value subtests of the Patterns of Adaptive Learning Scales (PALS) (Midgley et al., 2000) and Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich 1991) were used. Recent studies have shown that the MSLQ is the most widely used measure for SRL and self-efficacy (Roth et al., 2016; Honicke & Broadbent, 2016) however, PALS provides goal orientation measures specific to the theories of intelligence. The subtests from the PALS include: (1) mastery-goal orientation, (2) performance-approach goal orientation, and (3) performance-avoid goal orientation. The subtests from the MSLQ include: (1) task value, (2) control of learning beliefs (growth mindset), and (3) self-efficacy for learning and performance (see Appendix A). Each subtest consists of four to eight questions for a total of 26 questions. The statements were modified to specify “math class” instead of the generic “this course” used in the original questionnaire. The questions were presented in ASSISTments and students entered their response using either a five or a seven point Likert Scale where 5 and 7 indicated “very true” and 1 indicated “not true at all”. The same form was administered to students in both conditions.
Baseline Learning Task: This task was designed to provide a baseline measure of learning and students’ self-regulatory behavior. The assignment consisted of 12 total questions that required students to solve equations. There were four different types of equation questions (1) one-step equations with multiplication and division, (2) one-step equations with addition and subtraction, (3) one-step equations with fractional coefficients, and (4) two-step equations. There were three morphologically similar questions in a row for each of the types mentioned above making a total of 12 questions. The first question of each type serves as a pre-test that will later be used to measure learning.

Students were presented a math question in ASSISTments and were provided several options for help, if needed (see Figure 4.1). They could request a video, a worked example, or a hint by typing in a keyword. The video was an instructional video from YouTube on the specific math topic. The example was a worked example of a similar problem, which showed how to solve that type of problem from start to finish. The hint was in the form of a text message that prompted students on the first step. Students could also request the answer, which would provide the correct answer. Students must answer the question correctly in order to move on to the next question. Students could request more than one type of help if desired. So if the video support wasn’t helpful, they could still request an example or a hint. They could also request the same type of help more than once, although the information provided was the same. This is an important feature of the assignment design because it allows us to observe their self-regulatory behavior. Specifically, we can observe whether students shift their help-seeking behavior as needed in order to learn the content. Students were automatically given correctness feedback with every entered response so they could self-monitor their performance. They could continue to enter different answers until they answered correctly or they could request the answer in order to move on. The same form was administered to students in both conditions.
Solve the equation for x.

\[-3x = 15\]

If you would like help answering this problem, you can choose from the following options:

- Video - watch a video that will teach you how to solve these kinds of problems (Type "vid"
- Example - see a problem like this solved (Type "example"
- Hint - check out a hint to help you get starting (Type "hint")
- Answer - to see the answer to this problem click the "Show Answer" button.

Figure 4.3. Sample question from Baseline Learning Task showing the different forms of help available upon request.

Baseline Learning Quiz: This quiz consisted of four questions related to solving equations. There were four types of questions included that were morphologically similar to the questions in the Baseline Learning Task. Therefore one similar question from each type was used to create the Baseline Learning Quiz. This allows us to compare their performance on this quiz to the pretest items in the Baseline Learning Task to measure learning as a result of the learning task. Correctness feedback was not provided, instead, students were informed that their response was recorded. The same form was administered to students in both conditions.

Class Dojo Videos: Class Dojo is a product of Big Ideas (2018). The website offers several video series on a variety of topics. The series used in this study were empathy (control condition) and growth mindset (experimental condition). The videos use the cartoon character, Mojo, a green monster and his friends to discuss these topics. In the empathy series, the friends discuss their feelings during a variety of situations. In the growth mindset series, Mojo's friend helps him persevere through an initial failure by explaining that mistakes help us learn. Discussion questions are also included with each video to ensure that students are engaged while watching
the videos and are processing the target message. Individual videos range from three to seven minutes. Links to the appropriate videos were included in specific assignments. Participants in this study were already familiar with Class Dojo and the character, Mojo, as they used the behavior management piece of this program in class regularly.

**Teacher Created Videos:** In order to deliver very specific messages regarding task value and growth mindset, the teacher, who is also the researcher, created customized videos for each condition. The videos were inspired by Kelly et al. (2013). The videos used in the control condition provided content-specific math instruction delivered by the teacher. The videos used in the experimental condition displayed the teacher talking directly to the students. She explained that math homework is about so much more than practicing math, it help students to learn and provides an opportunity to practice persevering through challenges. These short videos were recorded using a cell phone and posted to YouTube. The links were included in specific assignments. The content of the videos varied by condition.

**Intervention Part 1 (Experimental Condition):** This assignment served as the start of the intervention. The assignment consisted of three videos and a series of open-ended questions related to the videos. The first two videos came from the Class Dojo series on growth-mindset. Each was followed with several discussion questions where students typed their responses. These questions served to ensure that students watched the videos, but also to force students to reflect on the critical messages of the videos. Questions included: “Do you think Mojo can become smarter? Why or why not?” and “What does the quote from the video, "mistakes make you smarter” mean to you?”. The third video was a video of their teacher delivering a message about the power of making mistakes. Following the video, students answered open-ended questions by typing their responses. Questions included: “What are some examples Mrs. Kelly gives of times when challenges and making mistakes can be fun and help us learn?” and “What advice would you give to someone who is getting frustrated on their homework because they keep making
mistakes?”. Based on prior successful interventions (Aronson, Fried, & Good, 2012), students were also asked to reflect on their personal experiences by answering questions such as, “Thinking about that time you learned from your mistakes, how did you feel when you were finally successful?” Only students in the experimental condition completed this assignment.

Control Part 1 (Control Condition): Serving as a control, this assignment was designed to be as similar as possible in structure to the assignment described above. Therefore, it also consisted of videos and a series of questions. The first video was the introductory video from the empathy series from Class Dojo. Students indicated that they watched the video and then answered several open-ended questions based on the video like, “What are some emotions that Mojo feels in the video? Why does he feel that way?” The second video was a video of their teacher delivering a lesson on combining like terms. Students then answered math questions practicing that skill. Only students in the control condition completed this assignment.

Intervention Part 2 (Experimental Condition): This assignment served as a continuation of the intervention. Students again watched three videos and responded to a series of open-ended questions. The first two videos were a continuation of the Class Dojo series on growth mindset. They specifically addressed brain function when learning and persevering through challenges. After each video, students responded to questions like, “Why is the word "yet" so powerful? Think of a "yet" moment in your life and write about it.” and “Sometimes when students are answering questions in ASSISTments they get frustrated because the questions are too challenging. Thinking about the video you just watched, what message can ASSISTments give to that student to help them get beyond their frustration?” Students then watched a video of their teacher who made the connection between the Class Dojo messages and challenging homework assignments. Specifically, that math homework is designed to be challenging to provide students the opportunity to make mistakes and therefore learn. Students were then required to reflect on that video and respond to the following question: “Based on what Mrs. Kelly shared, why
is learning math so important?” Only students in the experimental condition completed this assignment.

Control Part 2 (Control Condition): This assignment was designed to mirror the assignment described above. Therefore students were shown the next video in the empathy series from Class Dojo and responded to open-ended questions, such as, “What did MUM mean when they said ‘we can impact how others feel’?” Students then watched a video of their teacher modeling how to solve two-step equations. Students then answered a math question which required solving a two-step equation. Only students in the control condition completed this assignment.

Intervention Part 3 (Experimental Condition): This assignment continues to serve as part of the intervention. Students were first presented with a challenging math question that required them to simplify an algebraic expression. The purpose of this question was to present students with a struggle so that feeling was fresh in their mind as they engaged with the intervention. Students then watched the final video on growth mindset from Class Dojo. Again, they answered open-ended questions including, “What does it feel like when something is too challenging? What can you do when you feel that way?” Students were then shown a video from the Blackwell et al. (2007) study that presents the brain research supporting growth mindset. Students were then asked to directly connect the information they just heard to learning math with ASSISTments. To do this, they were asked, “The video says that you must work on challenging problems to grow your brain. However, you must use strategies to help you work through the struggle. What strategies can you use to help you when you are struggling with a challenging math question in ASSISTments?”

Next students were reminded that their grade is not based on their performance during the learning phases and that they should use the learning phase to prepare them for the assessment, which does determine their grade. They were presented with the following open-ended question,
“Remember, your grade in math class is not based on your ASSISTment score. Instead it’s based on you EVENTUALLY learning the material. That means, that you can get a question wrong, make many attempts, use hints, and even request the answer to help you figure out how to solve the problems. The important thing is that you use that to help you learn! How can doing homework in ASSISTments help you tackle challenging problems so that you can learn more math and grow your brain?”

The purpose of this question is to guide students to use the information they have recently learned through all of the videos to shift their behavior when learning using ASSISTments.

Students were then informed that the rest of this assignment is practice for their upcoming quiz. They were given an easy math question on simplifying algebraic expressions. Then they were asked if they wanted a more challenging problem to help them learn more or if they wanted another similar question (see Figure 4.4). This question serves as a measure of growth mindset rather than an actual part of the intervention.

**Figure 4.4. A question in the final intervention assignment designed to assess growth mindset in students.**

Regardless of the students’ selection, the assignment continued with nine additional challenging math questions that required students to combine like terms and simplify
algebraic expressions. Only students in the experimental condition completed this assignment.

Control Part 3 (Control Condition): This assignment is identical in structure to the assignment described above. The only difference is the content of the two videos. The first video was the final video of the empathy series from Class Dojo. Students were asked open-ended questions like, “How was Katie feeling before it was her moment to go out on stage? How do you know she felt this way?” They were then shown the control video used in Blackwell et al. (2007) that presents basic brain information not related to growth mindset. Students were asked to share, in writing, one thing they learned from the video.

The rest of the assignment was identical to the one presented above including informing them that the rest of this assignment is practice for their upcoming quiz. They were given the same easy math question on simplifying algebraic expressions, and asked if they wanted a more challenging problem to help them learn more or if they wanted another similar question. They were then given the same nine math questions to complete for practice. Only students in the control condition completed this assignment.

Intervention Part 4 (Experimental Condition): This assignment was designed to provide an opportunity for students to experience challenges while completing math work in ASSISTments and to apply what they learned from the previous assignments. Encouraging text messages based on performance were delivered throughout the assignment to reinforce the growth mindset messages presented in earlier intervention assignments and the final message included causal attribution features. See Table 4.1 for messages by performance and condition. The content of the messages were similar to the growth mindset condition in Arroyo et al. (2016).

This assignment begins with students answering two math questions that required them to add and subtract positive and negative mixed numbers. They were then given a third math question. If they answered it correctly, they were given an
encouraging message that praised their effort. Students who answered the question incorrectly, were given a text message that reminded them that it is okay that they made a mistake and to take advantage of the learning opportunity.

Students were then given two more math questions on the same topic for additional practice. They were then given one final math question. If they answered it correctly, they were again given an encouraging message that attributed their success to effort and using the presented challenge to help their brain grow. Students who answered that question incorrectly also received a text message that reminded them their mistake was beneficial to their learning.

Lastly, students were asked two questions about their level of frustration. First, they were asked how frustrated they felt right now. Then they were asked how frustrated they usually feel when doing math work. They responded using a 4-point Likert scale where 1 is Extremely frustrated and 4 is Not frustrated at all. Only students in the experimental condition completed this assignment.

Control Part 4 (Control condition): This assignment was designed to be identical in structure and mathematical content to the assignment described above. The content questions were the same. The only thing that differed in this assignment was the content of the text messages (see Table 4.1). Instead of reinforcing a growth mindset, they intentionally focused on performance, either correct or incorrect. Only students in the control condition completed this assignment.

Post-Test Learning Task: This learning task is designed using the same structure as the Baseline Learning Task that was completed at the start of the study. The assignment consisted of a total of 12 questions. Once again there were four topics included in the assignment (1) using equations to solve word consecutive integer problems, (2) solving multi-step equations, (3) given volume of a rectangular prism, find the height, and (4) using equations to find missing angles. There are three morphologically similar questions for each of these topics, resulting in 12 total questions.
Each question offered students several options for help. They could select a video, example, or hint by typing in the key word. They could also request the answer if needed. Students could request more than one type of help, if needed, and they could request the same type more than once. This allowed us to observe the self-regulatory behavior of self-monitoring and help-seeking. Correctness feedback was provided with each entered response. A correct answer was required to move on to the next question. Students in both conditions completed this assignment.

Table 4.1. Text messages embedded in the assignment to reinforce specific attributes depending on condition and performance.

<table>
<thead>
<tr>
<th></th>
<th>Experimental Condition</th>
<th>Control Condition</th>
</tr>
</thead>
</table>
| **Initial Message** | **Correct Response**
Great effort on a tough question. Remember, if you have trouble, it’s ok! Challenges help your brain grow stronger! | Looks like you are doing great. |
|                  | **Incorrect Response**
Remember, challenges help to make your brain grow stronger. Learning happens when you make a mistake. So take some time to try and figure out what you did wrong and then learn from that mistake. You can do this! | It looks like you are having trouble. |
| **Final Message** | **Correct Response**
Congratulations! You worked really hard to master problems with fractions. Thanks to the challenge and learning from mistakes, your brain is now stronger and more prepared to learn new things in the future! | Congratulations, you finished the assignment. There are just two more questions for you to answer. |
|                  | **Incorrect Response**
It’s ok that you made a mistake on that question. Remember we need to challenge ourselves if we want to grow more neurons. Every mistake is an opportunity to learn. | Congratulations, you finished the assignment. There are just two more questions for you to answer. |
Post-Test Learning Quiz: This quiz was comprised of four questions. There was one question from each of the topics listed in the Post-Test Learning Task in order to assess learning. Students were not given correctness feedback. Instead, they were told their answers were recorded.

4.2.3 Design This randomized controlled trial was conducted over the course of six 50-minute sessions. It was administered during students’ regular math class. Students were blocked on prior knowledge and overall hint use in ASSISTments before the start of the study. They were then randomly assigned to one of two conditions, control (n = 43) or experimental (n = 43). Students remained in the same condition throughout the entire study. The entire design by condition is presented in Table 4.2.

4.2.4 Dependent Measures

The interventions in this study were designed to directly impact students' self-beliefs. The hypothesis is that shifts in self-motivational beliefs will impact student behavior while learning, which will lead to increased learning. This section describes how each of these constructs will be measured along with the hypothesized effect.

Student Beliefs: The Self-Motivation Belief Survey was used to measure student beliefs. It was administered at the beginning and end of the study. Shifts in students’ responses for each subtest will be examined. It is hypothesized that growth mindset, mastery goal-orientation, task value, and self-efficacy will increase while performance-approach goal orientation and performance-avoid goal orientation will decrease.
### Table 4.2 The experimental design by condition. Shared assignments and differences between conditions are highlighted.

<table>
<thead>
<tr>
<th>Day</th>
<th>Experimental Condition</th>
<th>Control Condition</th>
</tr>
</thead>
</table>
| 1- Initial Data Collection | • Self-Motivation Belief Survey  
• Baseline Learning Task  
• Baseline Learning Quiz |                                                                                 |
| 2- Intervention Part 1 | • Growth mindset videos (Class Dojo)  
• Discussion questions  
• Teacher motivational video  
• Reflection questions | • Empathy Video (Class Dojo)  
• Discussion questions  
• Teacher instructional video  
• Math questions |
| 3- Intervention Part 2 | • Growth mindset videos (Class Dojo)  
• Discussion questions  
• Teacher task value video  
• Reflection questions | • Empathy video (Class Dojo)  
• Discussion questions  
• Teacher instructional video  
• Math question |
| 4- Intervention Part 3 | • Growth mindset video (Class Dojo)  
• Discussion questions  
• Brain research on growth mindset video  
• Reflection questions  
• Challenge question  
• Math practice | • Empathy video (Class Dojo)  
• Discussion questions  
• General brain function video  
• Comprehension question  
• Challenge question  
• Math practice |
| 5- Intervention Part 4 | • Math practice  
• Text messages targeting growth mindset  
• Frustration Survey | • Math practice  
• Text messages targeting performance  
• Frustration Survey |
| 6- Post-test | • Post-Test Learning Task  
• Post-Test Learning Quiz  
• Self-Motivation Belief Survey |                                                                                  |

**Proximal Outcomes (Behavior):** There are several specific self-regulated behaviors as well as other behaviors that are considered desirable while completing math assignments in ASSISTments. It is hypothesized that these behaviors will improve based on students’ shifts in beliefs. Using Post-Test Learning Task as a measure of the impact of the intervention, differences in these specific behaviors will be compared between conditions. Each behavior is outlined here.
Time Spent After Help Request: The time between students requesting help (either a video, worked example, or hint) and their next action is logged. This time is a measure of help-seeking behavior because it indicates that students are using the help appropriately rather than “gaming the system” (Baker et al., 2004). If a student is effectively using the help strategy they requested, they should be spending time looking over that help and using it to attempt the question again. This takes time. Therefore, it is hypothesized that time spent after a help request will increase.

Time Spent After Error: The time between students entering an incorrect response and their next action is logged. According to the growth mindset literature, an error is an opportunity to learn. Therefore, students should be taking time after an incorrect response to uncover their mistake and to try the problem again. This takes time. Therefore, it is hypothesized that time spent after an error will increase.

Frustration: At the end of Intervention 5, students are asked to provide feedback on their level of frustration at that moment as well as in general. The difference between those reports by condition will be examined. Students who have a growth mindset, should not get as frustrated after mistakes as students who have a fixed mindset because they view those moments as a learning opportunity. Therefore it is hypothesized that frustration will be lower for students in the experimental condition than the control condition.

Challenging Question Choice: During Intervention Part 4, students are given the opportunity to request a more challenging question to help them learn or simply additional practice. Choosing a more challenging question has been used as a behavioral measure of growth mindset. Therefore it is hypothesized that students in the experimental condition will select the challenging question more often than students in the control condition.

Assignment Completion: The number of students who completed the assignment will be compared by condition. Students with a mastery goal-orientation and high self-efficacy tend to persist through challenges. Therefore it is hypothesized that completion rates will be higher for students in the experimental condition than the control condition.
Learning: Ultimately the goal is to increase learning as a result of these mindset and behavioral shifts. To measure changes in learning, a learning score for both the Baseline Learning Quiz and Post-Test Learning Quiz must be computed and compared. To compute learning, either at baseline or post-test, we must first compute initial knowledge. To do this we will use performance on the first question for each topic in the learning task. Remember, there were four topics within each learning task and three morphologically similar questions for each topic. Initial performance on the first question will be used to indicate prior knowledge. The post-test also consisted of four morphologically similar questions and can therefore be used to indicated post knowledge, or knowledge after the learning opportunity. Differences in prior and post knowledge will be considered learning. This can be done for both the Baseline Learning Quiz and the Post-Test Learning Quiz.

4.3 Results

4.3.1 Student Beliefs

Means and standard deviations for each subtest by condition were calculated (see Table 4.3). Students in both conditions were balanced on each subtest according to the baseline survey data. Both the PALS and MSQL report tool mean and standard deviations for each of their subtests. A single-sample t-test comparison of means was conducted for each subtest. Results indicate that participants at the start of the study reported significant differences from the tool means for all of the subtests except mastery goal-orientation ($t = 1.43, p = 0.16$). Participants in this study reported having a higher performance-approach goal orientation ($m = 2.85$) than the mean reported for the measure ($m = 2.46$) $t=3.26$, $p = 0.001$. Similarly, participants report having a higher goal orientation ($m = 3.16$) than the mean reported for the measure ($m = 2.4$) $t=6.4$, $p < 0.0001$. Participants reported having a lower task value ($m = 5.30$) than the mean reported for the measure ($m = 5.54$) $t=2.02$, $p = 0.04$. Participants reported having a lower growth mindset ($m = 5.04$) than the mean reported for the tool ($m = 5.74$) $t=6.85$, $p < 0.0001$. Similarly, participants reported
having a lower self-efficacy (m = 5.15) than the mean reported for the tool (m = 5.47) t=2.60, p = 0.01.

To measure change in beliefs, all of the mean subtest scores were added together to compute an overall survey score for both the Baseline Survey (m = 25.51) and Post-Test Survey (m = 25.78). A paired t-test revealed no significant change in beliefs overall (t (79) = 0.14, p = 0.8). Similarly, an analysis of individual subtests revealed no significant changes with one exception. An analysis of the performance-approach goal orientation post-test survey data revealed that students in the control condition reported having higher performance-approach goal orientation (m = 3.10) than students in the experimental condition (m = 2.6), however this difference was only marginally significant (t (79) = 1.52, p = 0.06). This finding suggests that students in the control condition are more concerned with performing well and ensuring that others know they performed well than students in the experimental condition. For example, students in the control condition agreed more strongly with the following statement than students in the experimental condition: “One of my main goals is to show others that I’m good at my math classwork.”

A further analysis of the performance-approach goal orientation subtest was warranted. Mean baseline survey scores were subtracted from mean post-test survey scores to calculate shifts in this belief. A t-test of these values reveals that students in the experimental condition decreased this orientation (m = -0.29) while students in the control condition increased their orientation (m = 0.09) (t (79) = 1.94, p = 0.05). Remember, that low performance-approach goal orientations are more desirable for SRL. This suggests that the intervention effectively shifted this belief.
Table 4.3 Student self-belief survey results by condition for pre- and post-intervention compared to the norm-referenced values. *Indicates a significant difference.

<table>
<thead>
<tr>
<th>Subset (Scale Range)</th>
<th>Condition</th>
<th>Baseline Survey Average (SD)</th>
<th>Post-Test Survey Average (SD)</th>
<th>Shift in Belief (Post-Pre)</th>
<th>Measurement Average Average (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mastery Goal Orientation (1-5)</td>
<td>Experimental</td>
<td>4.35 (0.70)</td>
<td>4.32 (0.66)</td>
<td>-0.03 (0.45)</td>
<td>4.15 (0.88)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>4.18 (0.66)</td>
<td>4.24 (0.70)</td>
<td>0.12 (0.55)</td>
<td></td>
</tr>
<tr>
<td>Performance- Approach Goal Orientation (1-5)</td>
<td>Experimental</td>
<td>2.85 (1.10)</td>
<td>2.60 (1.13)</td>
<td>-0.18 (0.99)</td>
<td>2.46 (1.15)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>2.85 (1.13)</td>
<td>3.01 (1.25)</td>
<td>0.09* (0.51)</td>
<td></td>
</tr>
<tr>
<td>Performance- Avoid Goal Orientation (1-5)</td>
<td>Experimental</td>
<td>3.11 (1.07)</td>
<td>3.18 (1.06)</td>
<td>0.29 (1.02)</td>
<td>2.40 (1.04)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>3.23 (1.12)</td>
<td>3.08 (1.17)</td>
<td>0.18 (0.86)</td>
<td></td>
</tr>
<tr>
<td>Task Value (1-7)</td>
<td>Experimental</td>
<td>5.28 (1.00)</td>
<td>5.28 (1.12)</td>
<td>0 (0.69)</td>
<td>5.54 (1.25)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.29 (1.14)</td>
<td>5.50 (1.08)</td>
<td>0.04 (0.77)</td>
<td></td>
</tr>
<tr>
<td>Growth Mindset (1-7)</td>
<td>Experimental</td>
<td>5.05 (1.01)</td>
<td>5.27 (0.98)</td>
<td>0.18 (1.02)</td>
<td>5.74 (0.98)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.05 (0.86)</td>
<td>5.06 (1.01)</td>
<td>-0.05 (0.84)</td>
<td></td>
</tr>
<tr>
<td>Self-Efficacy (1-7)</td>
<td>Experimental</td>
<td>5.06 (1.04)</td>
<td>4.91 (1.17)</td>
<td>-0.13 (0.88)</td>
<td>5.47 (1.14)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>5.21 (1.24)</td>
<td>5.19 (1.11)</td>
<td>-0.12 (0.76)</td>
<td></td>
</tr>
<tr>
<td>Overall (6-36)</td>
<td>Experimental</td>
<td>25.55 (4.47)</td>
<td>25.56 (4.02)</td>
<td>0.01</td>
<td>25.76</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td>25.48 (4.50)</td>
<td>25.98 (4.43)</td>
<td>0.5</td>
<td></td>
</tr>
</tbody>
</table>
4.3.2 Self-Regulated Behavior

Despite failing to detect overwhelming shifts in student beliefs, it is still important to analyze the potential impact of the intervention on student self-regulated behaviors.

**Time Spent After Help:** There were three types of help available to students on the Post-test Learning Task, video, example, and hint. Student data was broken into action level segments defined by a start and end action. E.g. the sequence {start – incorrect attempt – request hint – correct attempt} is broken into segments: {start – incorrect attempt}, {incorrect attempt – request hint}, {request hint – correct attempt} with the associated duration times. The duration of each segment starting with a help request (excluding requesting the answer) is then transformed as follows: Time is z-scored within help request type (i.e. video, hint, example). Modified z-score that uses median instead of mean and median absolute deviation instead of standard deviation were used to account for likely outliers that occur when logging time (Iglewicz & Hoaglin, 1993). Again, to account for outliers, Z-scored time is winsorized to three standard deviations (if the z-score is greater than 3, it is set equal to 3). Winsorized z-scored time is z-scored again within the problem to account for varying levels of difficulty between questions. Using modified z-scores, the transformed time (which accounts for differences caused by help type and problem) is then compared across conditions using a t-test. Students in the experimental condition (m = 0.53) spent significantly more time after requesting help before their next action than students in the control condition (m = -0.08) (p = 0.03). This suggests that the intervention had a positive impact on this behavior.

Additional analyses were performed to better understand the use of help within this assignment. An analysis was conducted to determine if there was a difference in the number of distinct types of help requested within each question. Remember, students could request multiple types of help or the same type of help multiple times within a question. Let’s consider the following student who completes three problems. On the
first, the student asks for a video, then a video again, then a hint. On the second, the student asks for an example. Then on the third, the student responds without requesting help. The number of distinct types of help requested would then be 2 on the first problem, 1 on the second problem, and 0 on the third. Therefore, the average number of different help types requested by the student was \( \frac{2 + 1 + 0}{3} = 1 \). This average was computed for each student on the Post-test Learning Task. Students in the control condition requested an average of 1.1 different types per problem while students in the treatment condition requested 1.2 different types per problem. While this suggests that the intervention encouraged students to seek more help before requesting the correct answer, this difference is not significant \( (p = 0.26) \).

**Time Spent After Error:** Using the same process described above time spent after an error before the next action was z-scored for each question to account for differences in difficulty. These z-scores were modified z-scores (using median and median absolute deviation) and were winsorized to three standard deviations. The transformed time was compared across conditions using a t-test. Results indicate no significant difference between time spent after making an error before the next action between the experimental condition \( (m = 0.53) \) and control condition \( (m = 0.66) \) \( (p = 0.5) \).

**Frustration:** For this analysis, only students who completed the assignment containing this measure were included (experimental condition \( n=29 \), control condition \( n=25 \)). This introduces a bias in the results due to a selection effect. Students with higher levels of frustration or lower knowledge may not have completed the assignment sufficiently enough to reach these survey questions and are therefore inadvertently excluded from the analysis. However, the completion rate does not differ significantly by condition (experimental = 67%, control = 63%), which means the bias would affect both conditions equally.

A t-test revealed that unexpectedly, students in the control condition \( (M = 3.08, SD = 1.04) \) actually reported less frustration during this specific assignment than students
in the experimental condition (M = 2.59, SD = 1.12) (t(52) = -1.76, p = 0.08). However this difference was only marginally significant. It is interesting to note that students in the control condition also reported that typically they have slightly less frustration in general when doing math work (M = 2.68, SD = 0.95) than students in the experimental condition (M = 2.28, SD = 0.92) however, this difference was also not significant (t(52) = -1.59, p = 0.12). These findings support the survey results and suggest that the intervention did not successfully shift self-motivation beliefs. See Table 4.4 for mean and standard deviations for general frustration and assignment specific frustration by condition.

Table 4.4. Mean (and standard deviations in parenthesis) Levels of Frustration Reported in General and During the Target Learning Assignment by Condition using a Scale of 1 to 4 where lower values indicate higher frustration.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Mean Frustration in General</th>
<th>Mean Frustration During Target Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>2.68 (0.95)</td>
<td>3.08 (1.04)</td>
</tr>
<tr>
<td>Experimental</td>
<td>2.28 (0.92)</td>
<td>2.59 (1.12)</td>
</tr>
</tbody>
</table>

Challenging Question Choice Only students who completed the target assignment with this measure were included in the analysis (n = 73). Seven students from the control condition were absent and therefore not included and eight students in the experimental condition were similarly excluded. 64% of students in the experimental condition requested the challenge problems, whereas only 51% of students in the control condition requested the challenge problems. A chi-square test of independence was performed to examine the relation between condition and problem preference. The relation between these variables was not significant, X2 (2, N = 73) = 1.17, p = 0.28. This suggests that the intervention may have a small effect but confirms the survey findings that there were no significant differences in growth mindset between conditions.

Assignment Completion: Only 18% of all students completed the assignment during the class period provided indicating that the assignment was more challenging than expected. As a result, completion rate seems to measure academic ability or gaming
behavior (using bottom-out hints throughout the assignment) rather than perseverance as intended. Nonetheless, differences in completion rates between the control condition (15%) and experimental condition (20%) were analyzed. A chi-square test of independence was performed to examine the relationship between completion and condition. Results indicate that there are no differences between conditions, $X^2 (2, N = 83) = 0.49, p = 0.48$.

Because completion rates were so low, the number of questions answered were analyzed instead. Students in the experimental condition ($m = 6.1$, $SD = 3.9$) completed slightly more questions than students in the control condition ($m = 5.4$, $SD = 4.1$) however, a $t$-test revealed that this difference is not significant ($t(81) = 0.81$, $p = 0.42$). Again, it is important to note that because time was fixed, the number of completed problems is not a valid measure of perseverance as intended.

### 4.3.3 Learning

Learning gains were measured at baseline and at the conclusion of the study (see Table 4.5 for mean learning gains and standard deviations by condition). Learning gains were not significantly different at baseline, indicating that the groups were balanced. Due to the unexpected difficulty and lack of completion of the post-test learning assignment, it is not surprising that learning on that task is lower than baseline. A paired $t$-test revealed that students in the control condition had significantly lower learning at post-test ($m = 0\%$) than at baseline ($m = 19\%$) ($t (28) = 3.43$, $p = 0.001$). Cohen’s effect size value ($d = 0.7$) suggested a moderate to high practical significance.

As expected, students in the experimental condition learned slightly less on the post-test learning quiz ($m = 5\%$) than on the baseline learning quiz ($m = 16\%$). However, a paired $t$-test indicates that this difference was not significant ($t (39) = 1.35$, $p = 0.18$). Cohen’s effect size value ($d = 0.3$) suggests a small effect that is not significant. This
suggests that learning was positively affected by the intervention, as the decrease in learning was not as severe as students in the control condition.

Table 4.5. Changes in mean learning gains between baseline and post-test by condition

<table>
<thead>
<tr>
<th>Condition</th>
<th>Baseline Learning Gains Mean (Standard Deviation)</th>
<th>Post-test Learning Gains Mean (Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>19% (25)</td>
<td>1% (31)</td>
</tr>
<tr>
<td>Experimental</td>
<td>16% (30)</td>
<td>5% (31)</td>
</tr>
</tbody>
</table>

Another way to think about these learning results is to examine how many students gained, lost or stayed the same when comparing learning gains at baseline to learning gains at post-test. For example, a student who had a learning gain of -25% at baseline but 50% at post-test would have gained, whereas a student with a learning gain of 0% at baseline and -25% at post-test would have lost. Table 4.9 shows that more students in the experimental condition (36%) gained than in the control condition (18%). This again suggests that the intervention may have had small a positive impact on learning.

Table 4.6. The number of students (and percentages) by condition whose learning gains between baseline and post-test “gained”, “lost”, or “stayed the same”

<table>
<thead>
<tr>
<th>Learning</th>
<th>Control Condition (N =28)</th>
<th>Experimental Condition (N=39)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gained</td>
<td>5 (18%)</td>
<td>14 (36%)</td>
</tr>
<tr>
<td>Lost</td>
<td>16 (57%)</td>
<td>20 (51%)</td>
</tr>
<tr>
<td>Stayed Same</td>
<td>7 (25%)</td>
<td>5 (13%)</td>
</tr>
</tbody>
</table>

However, these results should be interpreted with some skepticism, as there is a significant amount of missing data and potentially a selection effect. To account for missing data in the pre and post-tests, scores were calculated based on the questions completed. For example, if a student answered the first question incorrectly, the second question correctly and then skipped the third and fourth
questions, they would receive a score of 50%. Learning gains can only be calculated for students who have both a pre and post-test score. At baseline 97% of students (n = 83) have a learning score. However, at post-test only 80% of students (n = 68) have a learning score. To calculate learning gains, students must have a learning score at both baseline and at post-test (n = 67), which limits the analysis to 79% of the participants, which is not split evenly between the conditions. This analysis is based on 28 students in the control condition and 39 students in the experimental condition. There is not a well-defined explanation for the difference in these rates, however, it is important to note the potential selection bias.

**4.4 Discussion**

This study was designed to incorporate multiple elements of interventions that have been found successful in prior research. Therefore it was hypothesized that the effects would be at least as substantial as that prior work. While there are some results that suggest that the intervention had a positive impact on self-regulated sub processes and learning, there is not clear evidence to deem this intervention successful.

The apparent failure of the intervention may be due to limitations and/or flaws in the measurement tools. The measures of self-motivational beliefs, at the heart of this research, are well vetted (Roth et al., 2016) and are supported by the analysis of student behavior. However, as Winne & Perry (2000) point out, while accurate, they can be considered measures of aptitude, meaning they may not be sensitive enough to capture temporal changes in the sub process. Additionally, the Post-test Learning Task and Post-test Learning Quiz were inadvertently more challenging than anticipated. This lead to incomplete assignments and therefore missing data, as well as creating a floor effect. With most students performing so poorly, it is hard to detect differences.
Another explanation for the lack of findings is potential flaws in the intervention design. Delivering the intervention over four consecutive school days rather than shorter segments across several weeks may have impacted its efficacy. It is likely that the videos lost their novelty and students became disengaged rendering the messages ineffective. Additionally, with so many components spread out over several days, many students missed at least one piece of the intervention, which would impact its efficacy. It is also possible that combining so many elements from other interventions actually decreases the effectiveness of each one instead of creating a cumulative effect. Perhaps students could not process all of the information and different messages again rendering the intervention ineffective. Lastly, the lack of learning may be the result of ineffective tutoring within the learning assignments rather than students’ failure to regulate their behavior.

There is a disconnect between the growth mindset messages embedded within the intervention and the correctness feedback delivered throughout the learning tasks. On one hand students are being told that progress and learning are the most important things but each assignment focuses on accuracy and correctness. This mixed message may confuse students leaving them feeling like the messages are merely lip service because ultimately every assignment comes down to the number of correct responses. To address this, future research should focus the feedback on effort and progress rather than performance.

It is also possible that shifting self-motivational beliefs, including mindset, may not be as easy as prior research suggests. Significant shifts may take months or even years to accomplish. We know that the participants in the study had lower self-motivation beliefs than the measurement tool averages. Perhaps there are other characteristics about this class that make the students unique with different needs than participants in prior research meaning they needed more robust interventions than what was provided. One thing is clear from this research, developing self-regulated learning is a challenging task.
References


5 Who Wants to Self-Monitor?

The series of randomized controlled trials presented in this chapter are exploratory in nature and were designed in response to the findings in the prior study (Chapter 4) that students’ self-regulated behaviors did not change in response to the automatic recording feature in ASSISTments.

**Role in Self-Regulated Learning:** According to Zimmerman’s self-regulation cycle (Zimmerman & Campillo, 2003), a growth mindset *indirectly* impacts learning, in part, by encouraging students to engage in self-monitoring. The present research limits its scope to self-monitoring (see Figure 5.1). Specifically, the study measures students’ preference towards correctness-feedback as a means of accurately self-monitoring progress. It then attempts to shift their preferences in favor of correctness-feedback. Simply put, is it possible to convince students that correctness-feedback is a valuable tool in learning so that when given the option, they choose to receive feedback?

![Figure 5.1 The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, metacognitive self-monitoring, is highlighted.](image-url)
5.1 Background

One fundamental characteristic that self-regulated learners have is the ability to seek out help when needed. However, prior to requesting help, students must recognize that they in fact need help. This requires self-observation, which typically relies heavily on metacognition, or one’s awareness of one’s own thought processes. This includes students’ awareness of their strengths and weaknesses, the demands of the learning task, their current understanding of the topic, and present progress throughout a learning task (Isaacson & Fujita, 2006). Pintrich et al. (2000) compares this process to a thermostat on a furnace. The thermostat monitors the temperature and when the temperature drops too low, the thermostat tells the furnace to turn on. Similarly, students must monitor their learning such that when they are struggling they regulate their behaviors, cognitive strategies, motivation or affect to increase learning (Isaacson & Fujita, 2006).

This awareness can be supported with internal (informal metacognitive self-monitoring) or external feedback on performance (Zimmerman 1998). Prior research indicates that many students struggle with metacognitive self-monitoring because they overestimate their ability and performance in the absence of feedback (Isaacson & Fujita, 2006). However, intelligent tutoring systems are extremely adept at providing external feedback on mastery with formal self-recording (Koedinger et al., 1997) and progress (Arroyo et al. 2007). Like the thermostat on the furnace, correctness feedback provides the trigger for the furnace to turn on. More complex thermostat monitoring systems, like the Nest, identify patterns and provide feedback on trends to help the furnace function more efficiently. Likewise, intelligent tutoring systems can track patterns in learning and provide feedback to learners displaying their trends to help them become more efficient learners (Arroyo et al., 2007).

Prior research (see Chapter 4) suggests that when provided external feedback on mastery in the form of correctness feedback, not all students use that information to regulate learning. Some students engage in “gaming” behaviors (Baker et al., 2004).
However, others may lack the strategies or motivation to use that feedback to regulate learning. In fact, it is hypothesized that not all learners even desire that feedback. Therefore, the current study attempts to determine the desirability of the self-recording feature, correctness feedback, while completing math homework in ASSISTments, an intelligent tutoring system. Further it seeks to shift students’ preferences to encourage all students to request correctness feedback as they complete their math homework.

5.2 Baseline Data Collection

5.2.1 Experimental Design

Participants were 92 students in the 7th grade who participated as part of their math class. Baseline data was initially collected to determine the percentage of students who desired correctness-feedback while completing their math homework using the ASSISTments program. To establish this baseline percentage, at the start of a homework assignment, students were informed that they would be given a choice for their homework. The wording was intentionally biased towards selecting feedback (see Figure 5.2). See Kelly (2018) to view the assignment as a student.

For tonight's homework, you have a choice!!!
Would you like to complete the assignment WITHOUT feedback? This means you will not be told if you are right or wrong. You can just enter your answers and move on.
OR
Would you like to complete the assignment WITH feedback? This means you will be told if you are right or wrong and can correct your incorrect mistakes.

Select one:
☐ I do NOT want feedback.
☐ I DO want feedback.

Submit Answer

Show answer

Figure 5.2. Instructions at the start of a homework assignment that presents students with the choice to receive correctness feedback or not.
5.2.2 Results

Only 67 out of the 92 students (73%) enrolled in the class started this assignment, therefore only their data is included in the summary here. Of the 67 students who attempted the assignment, 31 students or 46% requested feedback. An initial analysis reveals that there aren’t any differences in the ability levels of the students who selected feedback versus those who did not. Specifically, the mean overall performance for students who selected feedback was 66.1% (standard deviation 10.7%) and for students who did not select feedback was 65.8% (standard deviation 8.9%). This suggests that knowledge level was not a factor in requesting feedback or not.

These results establish baseline data that about half of the students in this particular class would rather complete homework assignments without correctness-feedback despite the tool being readily available.

5.3 Randomized Controlled Trial

5.3.1 Experimental Design

The same 92 students in the 7th grade math class mentioned above were included in this randomized controlled trial. A few days after the initial baseline data collection students were given another homework assignment with the same choice, to receive feedback or not, at the beginning of the assignment. It may be important to note that this was only the second time they were given this sort of choice throughout the year. Students’ choices were honored. This means that students who did not want feedback completed the entire assignment without any correctness-feedback. They submitted their answer and were told that their answer was recorded but there was no indication of the accuracy of that response (see Figure 5.3).
Students who requested feedback were told at the completion of each question if they were correct (see Figure 5.4) or incorrect (see Figure 5.5). If the answer was incorrect, students were given unlimited opportunities to self-correct. The final correct answer was available upon request and was required to progress to the next question. The last three questions of the assignment served as a post-test and therefore did not provide feedback to any student regardless of their original choice.

Once the post-test was analyzed, students were blocked on their choice and overall prior academic performance and were randomly assigned into two groups. Students in the control group (n = 43) were given the same choice of correctness-feedback or
not again (see Figure 5.1 for the precise wording). Students in the experimental condition \((n = 49)\) were shown a video before being offered the same choice of correctness-feedback or not (see Figure 5.6). See Kelly (2018) to view the assignments as a student.

Figure 5.6. The initial question in the homework assignment for students in the experimental condition. It includes the intervention, a video, and offers students a choice regarding whether they want to receive feedback or not.

This video intervention was designed with three key elements: (1) information was provided by an expert, (2) peer preferences were acknowledged to leverage peer pressure, and (3) students’ own performance was acknowledged to make the intervention relevant to students. The video starred professor Neil Heffernan, the creator of ASSISTments, because students are more willing to accept academic information from a perceived expert (Liu, 2004). The content of the video included two critical facts about student preferences and performance in order to educate and convince students to choose correctness feedback. First, students were informed that half of their classmates selected feedback on the prior homework assignment. This information was included to make students aware of their peer’s preferences, hoping to take advantage of the positive effects of peer influence (Shin, Daly, & Vera, 2007). Second, students were informed that their peers who selected feedback scored 30%
higher on the post-test than those who did not. This was included in an attempt to educate students about the benefits of correctness feedback on learning using data that was specifically connected to them. Students were then offered the choice to receive feedback or not while completing the assignment that followed.

5.3.2 Results

Students who failed to begin the assignment were excluded from the analysis. This means, in the control condition, 11 students (26%) were excluded for failing to begin the assignment. In the experimental condition 13 students (27%) were excluded for failing to begin the assignment. An additional 5 students (10%) in the experimental condition were excluded for failing to respond to the target question. That is to say that they began the assignment and saw the target question with the video but never answered that question or progressed further in the assignment. It appears that the video itself, or simply the presence of the video in the experimental condition may have deterred some students from continuing with the assignment and could be responsible for the differential drop out rate.

If you recall, the baseline data, presented in 5.2.2 above, suggests that 46% of students typically request correctness feedback. The findings of this randomized controlled trial are consistent with those findings. In the control condition, 14 students out of the 31 students (45%) who began the assignment requested feedback when given the choice. In the experimental condition, 17 students out of the 31 students (55%) who began the assignment requested feedback when given the same choice after being shown the video highlighting the benefit of feedback. This suggests that perhaps the video intervention may have successfully increased the number of students requesting feedback. However, a chi-square test revealed that this difference is not significant ($X^2 (2, N = 62) = 0.58, p = 0.4$). It is important to note, that the experimental condition had a disproportionate dropout rate. If we consider all students in both conditions, including those students who failed to start the assignment, these percentages drop significantly. In fact, 34.7% of all students in the
control condition requested feedback and 34.6% of all students in the experimental condition requested feedback.

A closer analysis of the data was needed to determine which, if any, students, changed their request for feedback from the initial data collection assignment to this assignment. To examine this, all possible combinations of responses were analyzed and are presented in Table 5.1. For the students who participated in both assignments (n=53), results indicate that overwhelmingly students do not change their preference for feedback. In fact, 72% of students made the same choice to receive feedback or not on both assignments. The intervention in this experiment was designed to convince students who had previously opted not to receive feedback on the first assignment to request feedback on the second assignment. We found that 10% of students did in fact make such a change in their preference. However this did not differ by condition. It is also important to note that 7% of students who previously requested feedback decided not to receive feedback on the second assignment. Again this did not differ by condition.

<table>
<thead>
<tr>
<th></th>
<th>Stayed the Same</th>
<th>Changed</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes Both Assignments</td>
<td>No First, Yes Second</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>7 (16%)</td>
<td>5 (12%)</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>11 (22%)</td>
<td>4 (8%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No Both Assignments</td>
<td>Yes First, No Second</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>11 (26%)</td>
<td>3 (7%)</td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>9 (18%)</td>
<td>3 (6%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Did Not Complete Both Assignments</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>17 (40%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>22 (45%)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
These results indicate that the video intervention was not effective in persuading students to select feedback while completing their math homework. In fact, the majority of students do not shift their preferences at all.

5.4 Follow Up Study and Survey

5.4.1 Experimental Design

Weeks after the randomized controlled trial presented above, another randomized controlled trial was conducted using a different intervention. The same 92 students in the 7th grade math class were randomly assigned, using the randomization tool in ASSISTments, to either the control condition ($n = 44$) or the experimental condition ($n = 38$). This assignment was completed during class instead of for homework.

Students in the control condition were again given the choice of receiving feedback or not. The initial question (see Figure 5.7) offering the choice was intentionally very leading and relied on growth mindset messaging (Dweck, CITATION) in an attempt to increase feedback selection. If students selected feedback, they were asked to explain why they chose feedback. Students who chose not to receive feedback were asked to explain why they did not want feedback. See Kelly (2018) to view the assignment as a student.

![Figure 5.7. The first question of the homework assignment which provides a choice to students whether they would like to receive feedback or not.](image)

For students in the experimental condition the first question was the same initial question regarding feedback as shown in Figure 5.7. Students who requested...
feedback were asked why they chose feedback, just like in the control condition and then proceeded to complete the assignment receiving feedback. Despite being in the experimental condition, these students did not receive the intervention. Students who selected that they did not want to receive feedback were asked why they selected that option. Then they were presented with the intervention (see Figure 5.8), which was simply the question, “How do you expect to fix your mistakes if you don’t know what they are?”

![How do you expect to fix your mistakes if you don’t know what they are?](image)

**Figure 5.8. The intervention, a text message, shown to students in the experimental condition who initially requested not to receive feedback designed to encourage students to select feedback.**

Would you like to receive correctness feedback while completing this assignment? This will allow you to learn from your mistakes to master this skill.

![Select one:](image)

**Figure 5.9. The question immediately following the intervention providing students another opportunity to request feedback.**

Remember, the purpose of the intervention is to encourage students who do *not* want correctness-feedback to select feedback. While the intervention relied on the growth mindset belief that mistakes are for learning, it was designed to be informative regarding the benefits of correctness-feedback. It was also intentionally worded to make it clear that the teacher wanted the students to request feedback, thus adding another layer of intentional persuasion. Once students selected the only possible
answer, “I understand that the purpose of this assignment is to help me learn how to solve inequalities”, students were then asked again if they would like correctness-feedback (see Figure 5.9).

5.4.2 Results

Sixteen students were excluded from the analysis because they were absent and never attempted the assignment. Out of the 82 students who participated in the study, 57 students (70%) requested correctness-feedback when initially asked. This did not differ by condition. Specifically, 68% of students in the control condition requested correctness-feedback while 71% of students in the experimental condition also requested correctness-feedback. It is important to note that these percentages are higher than both of the previous studies mentioned above in 5.2.2 and 5.3.2. One potential explanation for the increase in requests for correctness-feedback is that this assignment was completed in class rather than for homework. Student behavior might somehow be different during school hours versus at home. For example, students might expect to have to work harder or are more willing to expend effort in class whereas expectations around homework might be different. An alternative explanation is that the proximity of the teacher might have influenced the students to choose feedback when they might not otherwise.

The target population of this study is the students assigned to the experimental condition who initially declined to receive feedback (n=10). After receiving the intervention, and being asked again their preference for feedback, only one student (10%) changed his response. Note that this percentage is similar to the percentage of students found to change their preference in section 5.3.3. This clearly suggests that the intervention was not successful in convincing students to request feedback.

A deeper review of the survey questions is necessary to understand why some students are so reluctant to receive correctness-feedback. Remember, after students selected to receive feedback or not, they were asked to explain why they chose that
response. For students who requested correctness-feedback (n = 57), 13 students failed to answer the survey question and three students’ responses were incoherent or irrelevant. The remaining responses were very similar to each other, in fact, every other response mentioned wanting to know if they were right or wrong and/or appreciating being able to learn from their mistakes. To be more specific, 15 students spoke directly to wanting to know if they were right or wrong. Another 26 students said that they wanted that information to help them learn, improve and/or grow. See Figure 5.10 for sample student responses. (Responses are intentionally left exactly as written without correcting spelling or grammar.)

<table>
<thead>
<tr>
<th>Explanations for choosing feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>• “I chose to receive feedback because it helps me see whether I'm improving or not. :)”</td>
</tr>
<tr>
<td>• “I chose to receive feedback because I want to learn from my mistakes I make.”</td>
</tr>
<tr>
<td>• “I chose to receive feedback because I would want to know where I made a mistake rather than get the problem wrong. I would probably want to correct my answer if it was wrong so I don't make the same mistake.”</td>
</tr>
<tr>
<td>• “I like to get feedback so that if I'm wrong on a question I can try again to figure out the correct answer.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Explanations for choosing NO feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>• “because I don't like to get feedback and just do the problem without feeling nervous that I am going to get it wrong.”</td>
</tr>
<tr>
<td>• “I do not choose feedback because I get frustrated when I get the answers wrong and I just want to pull my hair out.”</td>
</tr>
<tr>
<td>• “because im ready to be right on my own”</td>
</tr>
<tr>
<td>• “it makes me frustrated when I keep getting the answer wrong”</td>
</tr>
<tr>
<td>• “I chose to not use feedback because sometimes when I use the feedback I get aggravated if I get the question wrong. It only tells me what I did wrong, unless I choose to use a hint. Which then means I will get the question even more wrong. It makes me not want to do my work and makes me more and more aggravated”</td>
</tr>
<tr>
<td>• “I choose not to get feedback because I don't want to have to keep fixing my mistakes. This way is easier then having to do it over again and simply get it done.”</td>
</tr>
</tbody>
</table>

Figure 5.10. Sample student responses (intentionally unedited) when asked to explain why they requested correctness-feedback or not.

For the students who chose not to receive feedback (n = 25), the responses for why they made that decision were more varied. Three students failed to answer the survey question and responses from five students were incoherent or unclear. One student stated she preferred no feedback because it saved her time when completing homework. Two students said they didn’t need the feedback and three said they
didn’t want to know if they were right or wrong. However, the majority of responses indicated that feedback induces anxiety and/or frustration.

5.5 Discussion and Future Work

The present research attempted to change students’ preferences for correctness-feedback by teaching them how important self-recording is for learning. The hypothesis was that students did not understand the role self-observation plays in learning. Therefore, the interventions used two different methods in an attempt to educate students on the benefits of correctness-feedback. The first used information presented in video form from an expert. The second delivered the information in text form and leveraged pressure from the teacher in an attempt to encourage students to select feedback. However, we discovered that students are not likely to change their preference despite our varied attempts.

The responses to the survey question may shed some light on the ineffectiveness of the included interventions as well as possible next steps. Clearly, the students who chose to receive feedback recognized the benefits of self-recording, specifically correctness-feedback, on overall learning. Those who did not want the feedback largely mentioned anxiety and frustration induced by incorrect responses. This suggests that regardless of the potential academic benefits, self-preservation wins. Even if something is good for us, if it is too painful we will not pursue it.

Remember, correctness-feedback is included as a feature in many intelligent tutoring systems because it is one method to help students accurately self-observe. Like the thermostat on the furnace, it alerts learners to the need to adjust their behavior to ensure learning, however students must have skills to act on that information. Here is where future successful interventions might lay. For example, automatically providing tutorials for students who are unsuccessful yet fail to shift their strategies might help them move beyond incorrect responses and learn. It would be necessary to determine the optimal number of failed attempts prior to forcing the tutorial. Assessing the effectiveness of individual tutorials would also be critical. It is worth
considering that students don’t know that it is necessary to shift their behavior after multiple failed attempts. Therefore, a successful intervention might be to simply prompt students to select a different strategy (ie. use the hints, check your notes, watch a video, e-mail your teacher, etc). If students do not have the skills associated with these different strategies, then direct instruction on alternate strategies might be a successful intervention.

This present research focuses on correctness-feedback as the mechanism for self-observation. The survey results indicate that this type of feedback induces anxiety in some students and could be damaging to self-efficacy. Therefore future attempts at encouraging self-recording must address this frustration. Future research may include any of the following potential interventions. Feedback timing could be adjusted so that correctness-feedback was provided sporadically so as not to overwhelm students with incorrect responses. Causal attribution interventions to change the reaction to incorrect responses could be explored. Or programs could provide adaptive assignments so that students experience more success than failure hence reducing frustration and building self-efficacy.

Other potentially successful interventions may rely on alternative mechanisms, rather than correctness-feedback, to assist students in self-monitoring. For example, Arroyo et al. (2007) were successful using graphs to help students track their progress and identify trends in their learning. Therefore, future research should consider interventions that use patterns and trends in effort, motivation and performance to help students regulate their learning behaviors. Such interventions would also assist with causal attributions and might help to develop self-efficacy as well. For example, using the logging capabilities of an intelligent tutoring system, the relationship between help-seeking behavior and performance could be tracked and reported to students allowing them to recognize that when they use hints they are more successful on the next question. This still allows students to observe their performance and regulate their behavior but the self-recording includes more information that correctness alone.
In conclusion, correctness-feedback can be a successful tool for self-observing for some students. However, for others it is either not sufficient for them to regulate their behavior or the impact on self-efficacy is too damaging to be helpful. Therefore future work on self-observing should consider these factors when attempting to design new interventions.

References


6 STOP! Preventing Hint Overuse to Increase Learning

The initial randomized controlled trial presented here was originally submitted to the Artificial Intelligence in Education Conference 2018. A follow-up study, addressing flaws in the methodological and statistical analysis was conducted and has been added to the original submission to create this chapter.

Role in Self-Regulated Learning: An integral component of educational software is to mimic human tutors by providing help, on demand, to students. However, using that help effectively is a self-regulatory skill that must be developed. This randomized controlled trial was conducted to determine if a very simple intervention can prevent hint overuse, and increase learning as a result.

Figure 6.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, help seeking behavior, is highlighted.
6.1 Background

Intelligent Tutoring Systems (ITS) are now widely used in K-12 education because they have been found to enhance learning. One critical feature of ITS is that they provide various levels of help to students in an attempt to increase learning. Researchers in the field of self-regulated learning have been addressing this idea of help-seeking behavior outside the context of computer-assisted instruction (Karabenick, 1998). Self-regulation involves knowing when, what type, and how to seek necessary help (Zimmerman and Campillo 2003; Nelson-Le Gall, 1987). So how can ITS support this self-regulatory skill? Prior research has found that computer-based learning programs that provide on-demand help in the form of scaffolding, tutoring, and feedback, have had a positive effect on learning (Aleven & Keodinger, 2000, Bartholome et al., 2006).

Scaffolding is a method used to deliver increasingly detailed support to learners while solving complex problems (Aleven et al., 2006). Usually the first hint is designed to help the student begin the problem. From there, additional hints are provided often creating a step-by-step worked example. The final hint, sometimes called the “bottom-out hint” typically provides the answer allowing students to complete problems that are above their ability, yet still learn from the process.

However, there is some evidence that not all students use the available help effectively which impacts learning (Aleven et al., 2003). There are two types of help misuse that appear most frequently, help avoidance (failing to use help when needed) and help abuse (using help when not needed) (Roll et al., 2011). Some students learn to take advantage of the help provided. They engage in “gaming behavior” which is defined by Baker et al. (2004) as “behavior aimed at obtaining correct answers and advancing within the tutoring curriculum by systematically taking advantage of regularities in the software’s feedback and help.” Gaming behavior is especially problematic because it has been found to reduce the effectiveness of the system by decreasing learning (Baker, Corbett & Koedinger 2004). This behavior is
so problematic that researchers have developed detectors to highlight the behavior such that it can be addressed through various interventions.

One common form of “gaming” is to abuse the hints provided by over using them (Aleven & Koedinger, 2000). In other words, students use hints that are not needed just to obtain the final answer without having to exert effort to answer the question. Intelligent tutoring systems have attempted to combat this behavior by designing features that prevent hint abuse, such as setting a threshold of incorrect responses before offering help as well as initiating delays between hint requests (Roll et al., 2011). However, these are merely strategies to prevent abuse, they do not necessarily develop the self-regulated sub-processes needed for proper help-seeking behavior in the future.

Therefore, researchers have attempted to teach students the necessary metacognitive skills to better use available help within the system. To do this, first we must understand what ideal behavior looks like. Attempts to understand and combat hint abuse have revealed the complexity of the problem. It appears that in certain circumstances a specific help-seeking behavior may be effective, while in others it may not. Intuitively we know that when a student does not need help but uses it anyway, learning is adversely affected; yet when faced with a challenging problem, requesting help leads to improved learning (Wood & Wood, 1999). Surprisingly, students who fail to use help can actually learn more than struggling students who request help perhaps due to the struggle involved in persisting (Roll et al. 2014). This suggests that ideal hinting behavior is difficult to define making it challenging to develop interventions leading to optimum use of help within an ITS.

Nonetheless, researchers have attempted to actually teach students proper help-seeking behavior. By analyzing prior data, researchers were able to develop production rules against which to evaluate student choices within an interactive learning environment (Aleven et al., 2004). The Help-Seeking Tutor Agent was created to identify poor help-seeking behaviors and to provide feedback to students. A more sophisticated version, including self-assessment and metacognitive
feedback, yielded comparable results. Improper help-seeking behavior was reduced, yet these skills did not transfer to other domains nor did the reduction in help abuse lead to improved domain learning (Roll et al. 2006). This suggests that it’s possible to reduce the appearance of “gaming” behaviors however this alone does not lead to improved learning.

How do we encourage students to use hints effectively? How do we increase student persistence in order to increase learning? While complex artificial intelligence based interventions have attempted to address these questions and have shown promise in decreasing help abuse, they have failed to increase learning. The current randomized controlled trial introduces a simple intervention in an attempt to decrease help abuse, in the form of overusing hints, in order to increase learning.

6.2 Study 1 Methods

6.2.1 Participants

Participants were 79 students in a 7th grade math class in an urban middle school. The students routinely use ASSISTments, an ITS, as a regular part of their math class for classwork and homework. Data retrieved from ASSISTments was used to create pairs of matched students. Specifically, students were matched based on their overall hint usage and current average in ASSISTments. Students within the pair were randomly assigned to either the control (n=39) or treatment condition (n=40).

6.2.2 Dependent Measures

Learning was computed separately for each of the three assignments throughout the study. Learning is measured by simply subtracting the pre-test score from the post-test score. This measure was not calculated for students who scored a 100% on the pre-test, as learning was not possible. Pre-test and post-test scores were computed by averaging student performance on two questions embedded within each assignment. Partial credit scoring was applied to both the pre- and post-tests. To
receive full credit for that item, a score of 100%, students must have answered the question correctly on the first attempt. If a student was able to answer the question correctly after one mistake without requesting a hint, they were given a partial credit score of 50% for that item. If they requested the answer or took multiple attempts to answer that item correctly they were given a score of 0%.

**Overall Hint Count** is measured by counting the number of hints requested throughout the assignment, not including the final hints that provide the answer. There were between three and five hints available per question for a range of 24-40 hints in total for each assignment.

**Bottom-out Hint Count** is measured by counting the number of times the students requested the final answer for a problem throughout the entire assignment. The maximum number of bottom-out hints for each assignment was eight.

### 6.2.3 Experimental Design

This randomized controlled trial was conducted over the course of three days (see Table 6.1). Each day, students completed one assignment, either as homework or in class. The Baseline Assignment was designed to collect baseline data on our dependent measures (student learning and hinting behavior) during a typical assignment. The next day, students completed the Intervention Assignment, which contained the intervention. Finally students completed the Post-Test Assignment, which included the delayed post-test items, as well as another opportunity to measure learning and student hinting behavior. See Kelly (2018) to experience the assignments as a student.

The structures of each assignment were intentionally similar with the notable exceptions of the content and the inclusion of the intervention for students in the experimental condition. Each assignment consisted of eight questions, which were divided into two quadruplets. All of the questions in each quadruplet were morphologically similar. The first three questions in the quadruplet provided hints
upon request to students. The hints were text showing each step of the problem stopping just shy of giving the final answer. The hints were designed to ensure that students could answer the question on their own. The final hint provided the answer, which is required for the student to move on to the next question. The fourth question in each quadruplet served as the post-test question and did not provide any hints but did allow students to request the final answer so they could move on to the next question.

Table 6.1. Study design showing timing and function of each experimental element

<table>
<thead>
<tr>
<th>Day</th>
<th>Assignment Name</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>Baseline Assignment (homework)</td>
<td>Baseline data on learning and hinting behavior</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Intervention Assignment (classwork)</td>
<td>Intervention</td>
</tr>
<tr>
<td>Thursday</td>
<td>Post-test Assignment (homework)</td>
<td>Delayed post-test (learning)</td>
</tr>
</tbody>
</table>

The structure of the Intervention Assignment was similar to the other assignments. For the first three questions within a quadruplet, students could request hints (See Figure 6.2 for a sample hint).

**Problem ID: PRABG33P**

**Solve:**

$$6 > \frac{7x + 9}{5}$$

First let’s multiply each side by 5.

$$5 \cdot 6 > \frac{7x + 9}{5} + \frac{5}{1}$$

**Fig. 6.2. Sample hints available upon request to students.**
Students in the control condition were then given the option to request the final answer. Students in the treatment condition were given one additional “hint,” which served as the intervention. The hint was a picture of the teacher holding a stop sign that asked students to read the message below the picture (see Figure 6.3). The message read, “Are you actually reading the hints? Please go back and look at the hints again. I KNOW that you will be able to answer this question on your own! Do NOT press hint again until you have read the hints and tried to answer the problem. Use these hints to help you learn how to solve this type of problem so that you can answer the next one correctly!”

Fig. 6.3. The intervention

While the text hints containing math support were identical between the control and treatment conditions, the number of actual hints was different because the intervention was included as a possible hint. As a result, we are unintentionally confounding number of hints with the actual intervention. When students in the control condition request a hint, they press a button that says “Show hint 1 of 2”, while students in the treatment condition see a button that says “Show hint 1 of 3”. This may alter hinting behavior.

The final assignment, Post-Test Assignment, followed the same structure as the other assignments. However, students were asked two additional math questions
with the same content as the Intervention Assignment to serve as a delayed post-test. Hints were not available however, correctness-feedback was provided and students could request the final answer.

### 6.3 Study 1 Results

Due to unusually high absenteeism, only 56 students out of the original 79 students (71%) were present for the Intervention Assignment. Completion of the homework assignments, Baseline (61%) and Post-Test (54%) was even lower. Included in the initial analysis were 29 students in the control condition and 27 students in the treatment condition. These students were still evenly matched across conditions as measured by overall hint count and performance prior to the start of the study.

#### 6.3.1 Baseline Data

An analysis of the Baseline Assignment revealed that, as expected, there was little difference between students in the two conditions. Performance, Overall Hint Count, and Bottom-out Hint Count were not significantly different (see Table 6.2).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Performance</th>
<th>Overall Hint Count</th>
<th>Bottom-out Hint Count</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>66% (25)</td>
<td>7.4 (7.0)</td>
<td>1.5 (2.1)</td>
<td>41% (36)</td>
</tr>
<tr>
<td>(N = 17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>68% (30)</td>
<td>7.7 (8.1)</td>
<td>1.8 (2.5)</td>
<td>19% (39)</td>
</tr>
<tr>
<td>(N = 18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Learning was not as balanced as we would have liked across the groups (41% verses 19%), however a t-test reveals that this difference was not statistically
significant \( (t(33) = 1.71, p=0.09) \). This assignment provides baseline data which can be used to compare hinting behavior and learning after the intervention.

### 6.3.2 Intervention Assignment

An initial analysis of the Intervention Assignment revealed that students were not evenly balanced on performance for this topic. Overall performance on the assignment was not significantly different \( (t(54) = -1.40, p = 0.17) \) however, students in the treatment condition scored higher \( (m=75\%) \) than students in the control condition \( (m=67\%) \). This difference was more pronounced on the first quadruplet (finding volume of triangular prisms). On the first question in the quadruplet, students in the treatment condition had a mean score of 52\%, while students in the control condition had a mean score of 38\%. This suggests that students in the treatment condition had more prior knowledge of this particular topic at the start of the study.

Yet despite this difference in initial knowledge, some differences in behavior and learning were observed (see Table 6.3). As expected, Overall Hint Count was slightly less for students in the treatment condition \( (m=2.8) \) than the control condition \( (m = 3.8) \) however, this difference was not significant \( (t(54) = 1.04, p = 0.3) \). Similarly, Bottom-out Hint Count was lower for students in the treatment condition \( (m = 0.4) \) than control condition \( (m = 1.4) \). An independent \( t \)-test revealed that this difference was significantly different \( (t(54) = 2.24, p = 0.29) \). Cohen's effect size was found to be moderate \( (d = 0.7) \). This indicates that while students in both conditions relied on hints, the intervention substantially reduced the frequency of requesting the final hint with the answer.

To determine the effect of this decreased hint abuse on learning, learning gains were compared. It is important to note that students with a pretest score of 100\% were not included in this analysis, as they couldn’t possibly learn. Furthermore, students in the treatment condition had higher levels of prior knowledge, therefore limiting the amount of potential growth as a result of the intervention. Nonetheless, students in
the treatment condition (m = 43%) had higher learning gains than students in the control condition (m = 39%). However, a t-test revealed that this difference is not significant (t(37) = 0.37, p = 0.7). Additionally Cohen’s effect size is minimal (d = 0.1) indicating that learning was not enhanced as a result of the decreased hint abuse.

Table 6.3. Means and standard deviations in parenthesis for Intervention Assignment by condition including performance as a percent, overall hint count, bottom-out hint count and learning as a percent increase

<table>
<thead>
<tr>
<th>Condition</th>
<th>Performance</th>
<th>Overall Hint Count</th>
<th>Bottom-out Hint Count</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>67% (24)</td>
<td>3.8 (3.8)</td>
<td>1.4 (1.8)</td>
<td>39% (29)</td>
</tr>
<tr>
<td>N = 29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>75% (15)</td>
<td>2.8 (3.0)</td>
<td>0.4 (0.9)</td>
<td>43% (35)</td>
</tr>
<tr>
<td>N = 27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.3.3 Baseline Assignment Target Population Only

In an attempt to determine the impact of the intervention, we re-analyzed the data using only a sub-set of students. This sub-set, our “target population,” included students in the treatment condition that saw the intervention, and the students in the control condition who would have seen the intervention if they were in the treatment condition. We found that only 7 students in the treatment condition saw the intervention while 17 students in the control condition requested enough hints that they would have seen the intervention had they been in the treatment condition. This difference in inclusion rates between the control and treatment condition is concerning. However, we argue that the difference in initial knowledge accounts for this differential inclusion rate. Specifically, students with higher initial knowledge require fewer hints and therefore reduced the number of students reaching the intervention.
Chapter 6 Hint Overuse

When re-analyzing the data from the Baseline Assignment, students in the “target population” were not very balanced (see table 6.4).

**Table 6.4. Means and standard deviations in parenthesis for the Baseline Assignment, for target population only, by condition including performance as a percent, overall hint count, bottom-out hint count and learning as a percent increase**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Performance</th>
<th>Overall Hint Count</th>
<th>Bottom-out Hint Count</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>59% (29)</td>
<td>9.7 (7.8)</td>
<td>2.4 (2.3)</td>
<td>35% (41)</td>
</tr>
<tr>
<td>N = 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>37% (31)</td>
<td>16.3 (8.0)</td>
<td>4.2 (3.0)</td>
<td>8% (38)</td>
</tr>
<tr>
<td>N = 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Students in the treatment condition actually used more hints (m = 16.3) than students in the control condition (m= 9.7). An independent t-test confirms that this difference is not significant (t(15) = 1.65, p = 0.1). Similarly, students in the treatment condition also used more bottom-out hints (m=4.2) than students in the control condition (m=2.4). Again, this difference was not significant (t(15) = 1.39, p = 0.2). Additionally, the students in the control condition appeared to learn more (m=35%) than students in the treatment condition (m = 8%). With such a small sample size of students who had a learning gain score, no significant differences were detected (t(14) = 1.29, p = 0.21).

### 6.3.4 Intervention Assignment Target Population Only

When analyzing the Intervention Assignment using only the target population, as expected the intervention successfully decreased Bottom-out Hint Count and increased Learning (see Table 6.5). Students in the treatment condition had a lower Bottom-out Hint Count (m=1.1) than students in the control condition (m=2.4). However, due to the small sample size, this difference was not significant (t(22) = 1.23, p=0.22). Cohen’s effect size suggests the effect of the intervention was
moderate (d = 0.78). Similarly, learning between students in the treatment condition was higher (m=57%) than students in the control condition (m=41%) yet this difference was not significant (t(19) = 1, p = 0.3). However, Cohen’s effect size is moderate (d = 0.46). This strongly suggests that the intervention may have had an impact on both behavior and therefore learning, but due to the small number of students included in the analysis we can not detect the difference.

**Table 6.5. Means and standard deviations in parenthesis for Intervention Assignment, for target population only, by condition including performance as a percent, overall hint count, bottom-out hint count and learning as a percent increase.**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Performance</th>
<th>Bottom-out Hint Count</th>
<th>Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>56% (25)</td>
<td>2.4 (1.8)</td>
<td>41% (35)</td>
</tr>
<tr>
<td>N = 17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>61% (18)</td>
<td>1.1 (1.5)</td>
<td>57% (35)</td>
</tr>
<tr>
<td>N = 7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A within-subject analysis comparing Bottom-out Hint Count from the Baseline Assignment to the Intervention Assignment revealed that for students who were shown the intervention, hint abuse was significantly decreased (t(5) = 3.16, p=0.02). Students in the treatment condition decreased their bottom-out hint use from a mean of 4.1 to 1.1 while students in the control condition did not decrease at all from a mean of 2.4 bottom-out hints. A paired t-test comparing student learning gains from Baseline Assignment to Intervention Assignment for students in the treatment condition revealed a significant difference (t(5) = 2.71, p=0.04). Specifically, students in the treatment condition increased their learning from a mean of 8% on the baseline assignment to 57% during the study. In comparison, students in the control condition only increased from 35% to 41%, which is not a significant improvement (t(7) = 1.76, p = 0.12). This suggests that for students using enough hints to trigger the intervention, the intervention was not only effective in decreasing Bottom-out Hint use
but also increased learning. Unfortunately, this learning did not persist as performance on the delayed post-test yielded no significant differences between conditions. See Table 6.6 for a summary of the within subject analysis.

**Table 6.6. Summary of Shifts in Bottom-out Hint count and Learning Gains between Baseline Assignment and Intervention Assignment for Target Population Only**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Baseline Bottom-Out Hint Count</th>
<th>Intervention Bottom-Out Hint Count</th>
<th>Baseline Learning Gains</th>
<th>Intervention Learning Gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>2.4 (2.3)</td>
<td>2.4 (1.8)</td>
<td>35% (41)</td>
<td>41% (35)</td>
</tr>
<tr>
<td>Experimental</td>
<td>4.2 (3.0)</td>
<td>1.1 (1.5)</td>
<td>8.3% (38)</td>
<td>57% (35)</td>
</tr>
</tbody>
</table>

### 6.3.5 Post-Test Assignment

An analysis of the second homework assignment, Post-Test, was completed only for the target participants (n=17). The purpose of this assignment was to detect if the effect of the intervention persisted beyond the Study Assignment. The analysis revealed that Bottom-Out Hint Count was still lower for the students in the treatment condition (m=1.2, sd = 1.8) than students in the control condition (m=1.9, sd=2.4), yet, not significantly different (t(15) = 0.60, p = 0.5). Cohen’s effect size suggests that this effect is in fact moderate (d = 0.3). A within-subject analysis of the five students who received the intervention and had data for both the Intervention Assignment and Post-Test Assignment confirms that the impact of the intervention on hinting behavior did not persist to the next day (t(4) = 0.93, p=0.4). However, that is largely due to one student who went from using zero bottom-out hints during the intervention to four hints on the following homework assignment. Unfortunately, learning could not be measured on this assignment because there was not enough data.
6.4 Study 2 Methods

Reasonable critiques of the prior study include the unequal number of hints provided to students by condition as well as the small number of students who received the intervention. Therefore, a follow-up study was conducted to address these flaws.

6.4.1 Participants

There were 75 students who participated in this study as part of their math class. They were the same students who participated in the study described above. Students were assigned to one of two conditions, treatment condition (N = 38) or control condition (N = 37). In order to assign students to conditions, they were blocked on prior academic performance in ASSISTments and overall hint use as well as prior condition assignment. Half of the students stayed in the same condition and half were switched to a new condition for this study.

6.4.2 Design

Unlike the prior study, this was conducted within one assignment during one class period. The assignment consisted of five parts. The first part served as the pre-test. There were two math questions, one on solving equations and the other on solving inequalities. There were no hints available however, students did receive correctness feedback and could self-correct. The correct answer was available upon request and was required to move on. The second part provided students directions for the remainder of the assignment (see Figure 6.5). They were told they would answer four practice math questions and that there were many hints available that they should use if needed. Students were required to agree to use the hints if needed. See Kelly (2018) to experience the assignment as a student.
Part three began the math practice and contained the intervention for those in the treatment condition. There were four practice problems that offered scaffolded hints. The number of hints varied by question but ranged from four to six. When students requested a hint, one step of the equation or inequality was solved for them. Each hint showed students the next step in solving the problem. For students in the treatment condition, the penultimate step was the intervention, which was the image of the teacher holding the stop sign urging students to read the hints before pressing hint again. To ensure that students in both conditions had the same number of hints available to them, the penultimate hint for students in the control condition restated the prior hint in different words.

Part four provided instructions about the post-test that was embedded in the assignment. Specifically it said, “The next 2 questions will not have any hints. Instead they will help test you to see how much you have learned. Do your best!” Part five, the post-test immediately followed. This section consisted of two questions that were morphologically similar to the questions in section one, the pretest. These were used to compute learning gains.

### 6.5 Study 2 Results

We focused on the four practice questions that provided hints. Then the number of of questions that students requested the bottom-out hints for (see Table 6.6). This value could range from zero to four. Students in the treatment condition on average used fewer bottom-out hints (m = 1.97) than students in the control condition (m = 2.19). However, an independent t-test revealed that this difference is not significant (t(73) =
Learning gains were calculated by subtracting students’ pretest score from their post-test score. This measure can only be calculated for students who completed both the pre- and post-test sections of the assignment. Additionally, students who scored a 100% on the pre-test were excluded from this analysis as they did not have any potential to gain. For most students, they performed worse on the post-test than the pretest resulting in negative learning gains. Students in both conditions had similar learning gains. This difference is not significant (t(56) = 0.08, p = 0.9). These results fail to replicate the findings of the prior study.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Average Bottom-out Hint Count (Standard Deviation)</th>
<th>Learning Gains (Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (N = 37)</td>
<td>2.19 (1.08)</td>
<td>-10% (33%)</td>
</tr>
<tr>
<td>Treatment (N = 38)</td>
<td>1.97 (1.15)</td>
<td>-11% (31%)</td>
</tr>
</tbody>
</table>

6.5 Discussion

In an attempt to increase learning, we introduced an intervention that was designed to support the self-regulated sub-process, help-seeking. While technology provides learners easy access to help while completing math work, using that help effectively requires self-regulation. Knowing when to request help and how much help to request relies on metacognition and is part of the self-regulating process. Describing desired help-seeking behavior is challenging because it may look different for different students or even in different situations. However, we acknowledge that there are certain behaviors that are almost always problematic. For example, students who are rapidly "clicking through hints" or simply relying on the final hint to provide the answer in order to move on are probably not using the help effectively.
Therefore, the studies presented in this chapter attempted to use a simple intervention in order to draw attention to hint abuse as a method of decreasing the overuse of help. For the purposes of this discussion, hint abuse refers to unnecessarily requesting the final hint which provides the answer. The assignment was constructed with hints that were intentionally designed such that the final hint should not be needed. Therefore, requesting the final answer suggests that students are gaming or lack the confidence in their ability to use the hints effectively.

We know there are many reasons that students choose to “game the system” and abuse hints. Once students begin hinting, some figure they have lost credit for the question anyway, so why not hint all the way through to the answer. Some students just hint their way through to complete the assignment quickly without having to do the work. Some students lack self-efficacy and don’t believe they can answer the questions correctly or learn from a computer and therefore rely on the final hints for the answers. We are not trying to understand the cause, yet the intervention was designed to potentially address several of these possible causes. For example, the shock of the picture was intended to stop students who were in the habit of hinting without thinking. Using the classroom teacher in the picture was intended to personalize the impact so that students were aware that the teacher cared, hopefully encouraging them to exert some effort. The message was designed to boost the confidence of struggling students and to encourage them to try the problem anyway. The hints were designed so that any student would be able to answer the question without having to request the answer. This was done to increase self-efficacy. If students experienced success, perhaps they would be encouraged to use the hints effectively in the future without abusing them.

While prior interventions from Aleven et al. (2004) relied heavily on artificial intelligence, this intervention was intentionally simple. We attempted to design an easy-to-create intervention that decreases hint abuse while actually increasing learning. This has yet to be accomplished with the more complicated interventions.
Unfortunately, fewer students in the treatment condition were exposed to the intervention than expected, resulting in a differential inclusion rate. While this difference may be due to differences in prior knowledge on one of the assignments between conditions, we acknowledge the challenge this posed to our initial analysis. Additionally, the number of hints shown on the screen was different for each condition, which, could have lead to some unaccounted for differences in behavior. We were able to address these issues in the follow-up study by ensuring an equal number of hints for both conditions and using more challenging content that encouraged students to use more hints as they completed the problems.

Initial results, from the first randomized controlled trial, indicated that the intervention may be beneficial in that we were able to decrease bottom-out hint count and increase learning gains. However, these results were not replicated in the follow-up study, which begs the question, what changed?

It is possible that during the replication study, the novelty of the intervention wore off and was therefore less effective. The differences in math content between the assignments may have had an impact. Considering the interconnected nature of the self-regulated learning sub-processes, it’s possible that for topics that are perceived to be more challenging, students have less motivation to persevere and might be more inclined to misuse help. Somewhat related to this is the possibility that our assumption that reaching a bottom-out hint is flawed. We must consider the role the hint is playing in the learning process. If a student needs help to start the problem then, continuing with hints is probably not beneficial to their learning. They may see the first hint and then continue the work on their own. However, a “low knowledge” student needs significantly more help. This means they may be using the hints, through the final answer, as a worked example. Even though they may be able to answer the question on their own, based on the penultimate hint, if they are using the hints as an example, it makes sense that they request all of the hints in order to see the full-example. They are likely still engaged and learning but to suddenly shift their thought process from observer to performer may not be natural or even helpful.
Furthermore, there is evidence that suggests the type of scaffolding used in this study, “fixed scaffolds”, provides only minimal learning gains even without hint misuse (Azevedo, 2008). Adaptive scaffolds that are informed by user learning characteristics may produce more significant learning gains (Greene, Moos & Azevedo, 2011). This suggests that despite more effective use of the hints by students in the treatment condition, the nature of the hints is to blame for the lack of increased learning gains. Therefore future attempts at decreasing hinting misuse should consider using adaptive scaffolds that are tailored to the needs of the learner rather than generic step-by-step assistance.

The results of these studies indicate that help-seeking is actually a rather complex behavior and therefore a simple intervention may not be sufficient to support or develop the self-regulated learning sub-process. Attempts at encouraging specific behaviors, like not using bottom-out hints, miss the nuances of self-regulated learning.

**References**


7 Addressing Hint Underuse with Choice of an Academic Tutorial

The study presented here was pre-registered at Open Science Foundation. At the time of submitting this dissertation, it has not yet been submitted for publication.

**Role in Self-Regulation:** The prior study focused on hint overuse, however, hint underuse also demonstrates weaknesses in self-regulation. Students are often reluctant to use the available help. It is hypothesized that strong performance-approach and performance-avoid goal orientations may be to blame. In order to make help appear more appealing, students are offered hints in the form of a tutorial. The present study attempts to increase persistence and learning during an assignment by offering students the choice to receive a tutorial before beginning the math assignment in an intelligent tutoring system.

![Cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, help seeking behavior, is highlighted.](image)

Figure 7.1. The cyclical model of self-regulated learning as presented by Zimmerman (Zimmerman & Campillo, 2003). The focus of this research, help seeking behavior, is highlighted.
7.1 Background

Many online math-learning environments require students to demonstrate proficiency in a skill by answering a certain number of consecutive questions correctly. ASSISTments is one example of an intelligent tutoring system that uses such a method to differentiate the amount of practice provided to students (Heffernan & Heffernan, 2014). Differentiation occurs because students who are already proficient in that skill require minimal practice and therefore, complete the assignment as soon as they demonstrate mastery. Students who understand the concept but struggle with fluency are provided just the right level of help and practice to complete the assignment. Lastly, students who have yet to learn the skill can access help in order to increase understanding and then practice to increase performance.

However, completing an assignment that requires a threshold of correct responses necessitates strong self-regulated learning skills. Ultimately the student must obtain a minimal level of knowledge within the assignment to reach such a threshold. Self-Regulated Learning (SRL) involves multiple sub-processes that work together to enable a student to control their own behaviors in order to learn (Zimmerman & Campillo, 2003; Panadero, 2017; Schunk & Zimmerman, 2008). These sub-processes, which include goal setting, self-motivational beliefs, help-seeking behavior, self-observation, and self-reflection are connected and together create a cyclical model of learning. In order for a “low knowledge” student to complete a math assignment, like the one described above, a student must be able to effectively access and use the help available within the intelligent tutoring system. In SRL, this is referred to as help-seeking behavior, which requires students to recognize when they need help, identify which type of help to use, and evaluate the effectiveness of that help (Roll et al., 2014).

Intelligent Tutoring Systems (ITS) aid students in this process by providing correctness feedback, which helps students to accurately determine when help may be necessary (Kelly et al., 2013; Keodinger et al., 1997). These systems also make
help readily available, often in the form of hints or scaffolds. However, there is evidence that students often fail to use the available help effectively (Aleven et al., 2003). Common misuses of help when learning inside an intelligent tutoring system include hint abuse (overusing hints when not necessary) and hint avoid (failing to use hints when necessary) (Roll et al., 2011; Karabenick, 1998). Hint misuse is problematic because it has been found to negatively impact learning (Baker et al., 2004).

To understand why some students are poor help-seekers, let’s consider self-regulated learning theory. Students’ behaviors during the performance phase (see Figure 7.1), are influenced by their self-motivation beliefs (Zimmerman & Campillo, 2003). These beliefs include self-efficacy, goal orientation and mindset. Students, who lack self-efficacy, or the belief that they have the ability to learn, will be less inclined to engage in productive learning behaviors (Bandura, 1982). Similarly, students with fixed mindsets, or the belief that intelligence is fixed, tend to engage help abuse (Dweck 2008). Additionally, students with high performance-approach or high-avoid goal orientations are likely to avoid using help. For example, imagine a student who is very concerned about appearing smart (performance-approach goal orientation) or fears a failing grade (performance-avoid goal orientation). She will avoid seeking help inside ITS, like ASSISTments, because her score will be lowered as a result, even though that help is needed to learn, complete the assignment, and ultimately achieve a higher grade.

As a result, the current research incorporates an intervention that provides students the choice to receive a tutorial before beginning an assignment. The tutorial prompts students to use the hints, including the final hint that provides the correct answer, in order to learn how to solve specific types of problems. This both gives permission to and actually encourages students to use the help provided and removes the focus on the assignment grade. It also helps students recognize that sometimes you need to learn a skill before you are ready to begin practicing it. And taking the time, in the beginning, to make mistakes and learn, may actually save time overall in completing
this type of assignment. Additionally, research suggests that just offering students a choice, can increase motivation (Aleven et al., 2004). Therefore, the intervention in the present study offers students a choice to receive the tutorial or not.

It is hypothesized that offering a tutorial to students at the beginning of a math assignment in ASSISTments will increase perseverance, as measured by assignment completion and learning. While this intervention could add time to the overall assignment length for students we predict that the benefit to student learning will be worth the cost in terms of extra time.

### 7.2 Experimental Design

#### 7.2.1 Participants

Students (n=82), in a 7th grade math class in an urban school district, participated as part of their math class. They routinely used ASSISTments, a computer-based learning program, to complete nightly homework and were therefore familiar with the platform.

#### 7.2.2 Materials

Materials used in this study include a pre-test, assignment, post-test, choice question, and a tutorial. See Kelly (2018) to experience each as a student.

The pre-test was given in class prior to the start of the study. It included questions on probability as well as statistics and was used as a classroom assessment for the course.

The assignment consists of three types of probability questions. (1) Finding the probability of two events without replacement. (2) Finding the probability of multiple independent events. (3) Finding the probability of the sum of two events. There are 84 randomly generated problems that are morphological copies of these types.
Students must continue to answer questions until they have correctly answered three consecutive questions, without requesting hints. Correctness feedback is provided automatically each time a student enters an answer. Hints are available upon request. Students must enter the correct answer to move on. If they can not find the answer themselves they can request the correct answer from the system.

The **post-test** consists of three total questions, one of each type mentioned above. Correctness feedback is provided each time a student enters an answer. Students have unlimited attempts to self-correct but must enter a correct answer before moving on. This can be obtained by requesting hints or the final answer.

The **choice question** is only available to students who are randomly assigned to the treatment condition (see Figure 7.2). Students are shown a sample question from the assignment and are asked if they would like a tutorial before they begin the assignment.

![Sample question](image)

**Figure 7.2.** The initial question for students in the treatment condition that offered students a choice to complete a tutorial or not before beginning the assignment.

Students, who at first decline the tutorial but go on to answer the next question incorrectly, are given a similar choice question asking if they would now like the
tutorial recognizing that they were unsuccessful on their first attempt (see Figure 7.3). This is a critical component to the design because prior research suggests that students often overestimate their abilities, so the automatic self-recording feature in ASSISTments provides the necessary feedback to help students recognize that they may need help (Aleven & Koedinger, 2000).

**Figure 7.3.** The question posed to students in the treatment condition who initially refused the tutorial but proceeded to incorrectly answer the first question in the assignment.

The tutorial is only given to students in the treatment question who requested help before beginning or continuing the assignment. The tutorial walks students through each of the three types of probability questions in the assignment by scaffolding the solution through a series of hints, with the last hint providing the correct answer. Once students enter the correct answer, they are presented with the next question. Students are not required to request every hint and they can enter the correct answer at any point during the tutorial to move on to the next question. Once the tutorial is complete, students are congratulated on completing the tutorial and are encouraged to begin the assignment with a reminder that the hints are available to help then whenever needed. It is important to note that students in the control condition also have access to the same questions and identical hints that comprise the intervention however they are not presented as a tutorial and must be requested individually. This ensures that learning gains are in fact due to the intervention and not enhanced instruction.
7.2.3 Procedure

Students were assigned one assignment on compound probability in ASSISTments. Using the randomization tool inside ASSISTments, students were randomly assigned to either the control or treatment condition (see Figure 7.4 for the experimental design). In the control condition, students were given the assignment and post-test. In the treatment condition, students were given the choice question. If they requested help, they were given the tutorial followed by the assignment and lastly the post-test. If they chose to begin the assignment without any help they were given one question. If they answered the question correctly, the assignment continued and was followed by the post-test. However, if they answered the first question incorrectly, they were given the second choice question. From here, if they requested help they were given the tutorial, followed by the assignment and lastly the post-test. If they chose to continue the assignment, then the assignment continued and was followed by the post-test. Students were given one week to complete their assignment. Students in both conditions who failed to complete the entire assignment after one week were assigned the post-test.

7.2.4 Dependent Measures

**Perseverance:** Perseverance is measured by assignment completion, which is a binary measure, complete or incomplete.

**Learning:** Learning will be measured using the post-test score. Post-test scores will be computed using a partial credit metric where the best score is a 1 and the worst score is a zero. Each hint requested deducts 0.25 credit and each incorrect response deducts 0.1 credit. The last hint that gives away the answer will always make the score for the problem 0. Here is one example: If there was a problem with 8 hints, then after 4 hints the score is zero. If the problem has only one hint and the student is told the final answer, their score drops to 0.75 with the first hint and then to zero with the second/final hint. Here is a second example: If a student asks for a hint and
then incorrectly attempts the answer followed by a correct response, the student will receive a $1-(.25+.1)=.65$ partial credit score. Once all hints have been used no credit is given. The partial credit scores for the three questions are averaged together.

![Diagram of experimental designs by condition. The treatment condition offers students a choice of a tutorial at the beginning of the assignment and again, a second time later, if the student did poorly the first question. The control condition does not include the choice or tutorial.]

**Figure 7.4.** The experimental designs by condition. The treatment condition offers students a choice of a tutorial at the beginning of the assignment and again, a second time later, if the student did poorly the first question. The control condition does not include the choice or tutorial.

**Time to completion:** The total time required to complete the assignment is computed within ASSISTments. After five minutes of inactivity time calculation for that action stops. It restarts with the student’s next action. Note that this dependent measure will only be calculated for those students who complete the entire assignment.
7.3. Results

Of the 82 students enrolled in the math class, only 72 participated in the study. Students were randomly assigned, using the randomization tool in ASSISTments, into one of two conditions, treatment (n = 38) or control (n = 34).

An analysis of the students in the treatment condition revealed that 26% (n=10) of students initially requested the tutorial. However, only 2 students who claimed they were ready to begin actually went on to answer the first question in the assignment correctly. This indicates that students over-estimated their ability. Of the 26 students who initially said they did not want the tutorial, when presented with the option again after answering the first question wrong, 12 (46%) chose to begin the tutorial. This means that out of the 38 students assigned to the treatment condition 56% (n = 22) received the tutorial (see Table 7.1). Four of those students failed to complete the tutorial and subsequently the rest of the assignment.

Table 7.1. The number of students in the treatment condition choosing to receive the tutorial when asked initially (first question choice) and after making an error (second question choice)

<table>
<thead>
<tr>
<th>Student Response</th>
<th>First Question Choice</th>
<th>Second Question Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>10 (26%)</td>
<td>12 (46%)</td>
</tr>
<tr>
<td>No</td>
<td>28 (74%)</td>
<td>14 (54%)</td>
</tr>
<tr>
<td>Total N</td>
<td>38</td>
<td>26</td>
</tr>
</tbody>
</table>

7.3.1 Completion

The number of students who completed the assignment (met the three consecutive correct questions threshold) and answered at least one of the post-test items were counted. A chi-square test revealed no significant difference in completion rate by condition ($X^2 = 1.27$, $p = 0.25$). In fact, students in the control condition had a slightly higher completion rate (82%) than students in the treatment condition (71%). Given
the relatively high percentage of students who requested the tutorial, this is a surprising finding.

Table 7.2. The number (and percentage) of students who completed the assignment by condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Complete</th>
<th>Incomplete</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (N = 34)</td>
<td>28 (82%)</td>
<td>6 (18%)</td>
</tr>
<tr>
<td>Treatment (N = 38)</td>
<td>27 (71%)</td>
<td>11 (29%)</td>
</tr>
</tbody>
</table>

7.3.2 Learning

Students in the treatment condition (m = 39%, SD = 33), scored slightly higher on the post-test than students in the control condition (m = 34%, SD = 27). However this difference was not significant, (t (1,68) = 0.64, p = 0.5) furthermore, Cohen’s effect size is small (d = 0.16). This suggests that the tutorial did not impact learning significantly.

7.3.3 Time to completion

The intervention takes a substantial amount of time to complete therefore, it is expected that students in the treatment condition would take longer to complete the assignment than students in the control condition. Results indicate that students in the treatment condition (M=24.6, SD=10.2) do not take significantly longer than students in the control condition (M=21.6, SD=9.2; t(1,50) = 0.95, p = 0.3). This finding is further supported by Cohen’s effect size (d= 0.3). In fact, students in the treatment condition took, on average, only three minutes longer to complete the assignment than students in the control condition. This suggests that the intervention may have reduced the time it took to reach the mastery threshold.

To further examine this claim, the number of questions it took students to reach the mastery threshold was compared by condition. On average, students in the
treatment condition completed 7.96 questions before reaching the three-correct-in-a-row threshold needed to complete the assignment, whereas students in the control condition completed 8.75 questions before reaching the same threshold. An independent t-test reveals that this difference is not significant ($t(1,51) = 0.66, p = 0.5$). Additionally, Cohen's effect size suggests that one question is a small effect anyway ($d = 0.18$).

### 7.4 Discussion

This randomized controlled trial was conducted in an attempt to prevent hint misuse. The focus of the study was on hint misuse in the form of avoiding hints even when they are necessary for learning. Using self-regulated learning theory to identify potential causes for the hint-avoiding behavior, lead us to create an intervention designed to encourage students to accept help in the form of a tutorial before beginning the assignment. About half of the students took advantage of this tutorial, but it is interesting to note that most of them did so only after learning that they answered a question incorrectly. This is further evidence that students do not always know when they need help without more formal self-recording strategies. Future attempts with a similar intervention should be sure to offer the tutorial at different points during the assignment, especially after students receive performance feedback. Conversely, it would be interesting to see if the tutorial is more or less effective when automatically delivered to students who have demonstrated a need based on performance.

Looking at all of the data from this randomized controlled trial together tells us that students in both conditions completed the same number of questions in the same amount of time despite the addition of the tutorial for students in the treatment condition. Additionally, students in both conditions experienced the same amount of learning. This suggests that while the tutorial did not increase completion rates or achievement, there is not a disadvantage in offering it. The lack of effect of the tutorial calls into question the quality of the help provided within the tutorial. Future
research should attempt this intervention again being certain to design a stronger tutorial.

This intervention was created to combat performance-approach and/or performance-avoid goal orientations. Perhaps the cause of hint misuse is another self-regulated learning sub-process such as task value, which affects motivation. Future research should focus on understanding why students are resistant to using help when needed and then build interventions that address that cause.

References


Self-Regulated Learning (SRL) has been found to be a critical component of learning and achievement (Pintrich & DeGroot, 1990; Zimmerman, 1990; McCombs, 1989). Therefore, in order to increase student performance, educational researchers have attempted to develop SRL skills such as mindset (Dweck & Leggett, 1988), help-seeking behavior (Aleven et al. 2004), and goal setting (Duckworth, 2015). These skills are especially important when learning is taking place while interacting within educational technology (Johnson & Davis, 2014). Fortunately, technology-based learning has the advantage of being able to support SRL skills and potentially develop them while also teaching content (Barak, 2010; Denton et al., 2008; Steffens, 2006).

The research presented here, based on both SRL theory and prior successful interventions, attempted to further understand and develop a variety of the sub-processes that make up SRL by delivering interventions through an intelligent tutoring system. Most of the studies were unsuccessful in shifting student beliefs or behaviors and therefore did not support prior research findings. This section will present some possible explanations as to why and provide considerations for future research.

There are a variety of models of SRL (Panadero, 2017). All of them are complex with interactions between each of the many sub-processes. Due to the theorized cyclical nature of SRL, developing one sub-process should have an impact on others as well (Zimmerman & Campillo, 2003). The key is understanding the relationships between the sub-processes and identifying where to start in the cycle. The present research attempted to leverage this cyclical aspect of SRL to impact learning. The failed attempts may suggest that the targeted sub-processes were not critical points in the cycle or the intended interactions were not significant enough. A meta-analysis by Dent and Koenka (2015) began investigating the relative importance of each sub-process as well as the correlations between them. Future research should continue...
to explore the relationships between the sub-processes in order to design relevant interventions that target the key features of SRL.

In addition to the complexity of SRL, students are also complex, each with unique needs. Students struggling with SRL do so for a wide variety of reasons. One student might have a fixed mindset while another may have a growth mindset but low perceived value of math, and another doesn’t know how to seek the appropriate level of help. Some students may also have multiple deficits in their SRL skills. While a few of the interventions included in the studies presented here were designed to address multiple sub-processes simultaneously, results indicate that this method may not be effective. It is possible that the students may have felt bombarded with multiple messages making the intervention too overwhelming to be effective. If this is the case, future research should consider adaptive interventions where the intelligent tutoring system identifies the weak SRL skill and delivers a targeted intervention.

One key finding from this research is that many students reported high levels of anxiety and/or frustration while learning math. According to the SRL cycle, this negative self-evaluation and self-reflection will impact all future learning experiences (Zimmerman & Campillo, 2003). Research in the field of psychology has confirmed that math anxiety is correlated with decreased performance (Wigfield & Meece, 1988). Therefore, future research should consider addressing this anxiety and frustration before targeting other SRL sub-processes. Prior attempts using intelligent tutoring systems have shown promise (Supekar et al., 2015; Fuchs et al., 2013). Perhaps this could also be done through adaptive assignments where the difficulty of questions is matched to the ability of the student, reducing the overall number of errors and increasing the overall feeling of success. Another option is to provide more support than correctness feedback or scaffolded questions with adaptive tutoring (Sottilare & Sinatra, 2014). This is tutoring that is specific to the error or needs of the student. Obviously, adaptive assignments and tutoring require substantially more artificial intelligence than the interventions discussed here, which may not be practical.
It is also important to recognize that the basis of SRL is learning as opposed to performance. Traditional K-12 education uses performance as a measure of learning. This places an emphasis on grades instead of growth. Many of the interventions were based on developing a growth mindset and emphasized the role of mistakes in the learning process, but ultimately the only feedback they received while practicing math was performance feedback. Therefore it could be argued that students were receiving mixed messages. Those responsible for educational policy are currently debating the impact of performance versus growth assessments. Data from both are both vital to assessing students and teachers and providing feedback related to SRL, however, they each play a different role (Lachlan-Hache & Castro, 2015). To address this, future research might explore different types of feedback, including growth and progress feedback to emphasize learning rather than performance.

Despite the hype in K-12 education about growth mindset, there is a growing body of literature that has failed to replicate the landmark findings presented in Dweck and Leggett (1988). Li and Bates (2017) failed on three separate accounts to replicate some key findings in the original growth mindset work. A meta-analysis of prior research also found that the relationship between growth mindset and academic performance is not as strong as initially reported (Sisk et al., 2018). That same meta-analysis discovered that growth mindset interventions did show some benefits to students from low-socioeconomic backgrounds and those in danger of failing but, even those effects were found to be smaller than originally anticipated. Another study reported that a growth-mindset intervention successfully shifted student beliefs but that shift did not lead to increase academic performance (Brougham & Kashubeck-West, 2018). The failed attempts to increase learning through growth mindset interventions presented throughout this dissertation may suggest that mindset does not impact learning or at least not through interactions with other SRL sub-processes.

Lastly, we should consider that perceived weaknesses in SRL skills are domain specific and therefore not something that need to be developed but rather transferred. For example, let’s compare the following students. Kim is an incredibly
successful student who demonstrates strong SRL in her academic classes. However, she struggles when playing video games and loses in the first round. She has a growth mindset and believes that if she practices enough she will be successful, yet she doesn’t care enough about video games to devote her time to learning how to progress through the levels and therefore, avoids them or fails miserable while playing. Dan is an avid gamer who demonstrates strong SRL when starting new video games. However, he struggles when learning math and often fails his tests. He has a growth mindset and believes that if he practices enough he will be successful, yet he doesn’t care enough about math to devote his time to learning algebra and geometry and therefore avoids math work or fails miserably. Both of these students have the ability to engage in SRL as demonstrated by their ability to independently learn and find success. However, their willingness, or lack there of, to apply these skills to certain domains determines their success. This suggests that interventions should focus on developing outcome expectations (the perceived purpose of the task) and task-value (the perceived value of the specific task) in order to motivate students to apply their SRL knowledge to academic areas. In addition, future research might attempt to leverage extrinsic motivation to encourage students to engage in SRL in multiple domains. It is also possible that students are just unaware that the skills that allow them to master video games or sports or music are the same skills that allow them to learn math. Therefore, other successful interventions might include explicitly teaching students how to transfer their SRL skills across domains.

As an educator, every day it is my job to educate all of my students. This includes the students who are eager to learn and easily engage in SRL as well as the students who struggle and resist learning. We need researchers to continue their work to understand the causes of these differences in motivation and success within students that have the same academic ability. Then we must develop successful interventions that address these differences so that each student reaches his or her full academic potential. The present research hypothesized that these differences stemmed from deficits in SRL. The results suggest that researchers should also consider other explanations for the differences in academic engagement and success.
References


You are about to complete a survey about yourself. Please answer the questions as honestly as possible. This is not a test and there are no right or wrong answers. Your answers will not affect your grade. Some questions might sound the same but it’s important to ask the question in different ways to make sure that your opinions are understood.

Some questions will use a scale from 1 to 5 where 1 means “not at all true” of you, and 5 means “very true” of you.

Other questions will use a scale from 1 to 7 where 1 means “not at all true” of you, and 7 means “very true” of you.

1. It’s important to me that I learn a lot of new concepts this year in math class.
2. If I study in appropriate ways, then I will be able to learn math.
3. I think I will be able to use what I learn in math class in other classes.
4. I believe I will receive an excellent grade in math class.
5. I’m certain I can understand the most difficult material presented in math class.
6. It’s important to me that other students in my class think I am good at my math class work.
7. It is my own fault if I don’t learn the material in math class.
8. It is important for me to learn the material taught in math class.
9. One of my goals is to show others that I’m good at my math class work.
10. I’m confident I can learn the basic concepts taught in math class.
11. One of my goals is to show others that math class work is easy for me.
12. It’s important to me that I look smart compared to others in my math class.
13. One of my goals in math class is to learn as much as I can.
14. I am very interested in math.
15. If I try hard enough, then I will understand math.
16. I’m confident I can do an excellent job on the assignments and tests in math class.
17. It’s important to me that I don’t look stupid in class.
18. One of my goals is to master a lot of new skills this year in math class.
19. I think the course material in math class is useful for me to learn.
20. It’s important to me that I thoroughly understand my class work in math.
21. If I don’t understand the course material, it is because I didn’t try hard enough.
22. It’s important to me that I improve my math skills this year.
23. Understanding the subject matter in math class is important to me.
24. One of my goals is to keep others from thinking I’m not smart in math class.
25. One of my goals is to look smart in comparison to other students in my math class.
26. Considering the difficulty of math, my teacher, and my skills, I think I will do well in math class.
27. It’s important to me that my teacher doesn’t think that I know less than others in my math class.
28. One of my goals in math class is to avoid looking like I have trouble doing the work.

<table>
<thead>
<tr>
<th>Sub-Tests</th>
<th>Item Numbers:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mastery Goal Orientation</td>
<td>1, 13, 18, 20, 22</td>
</tr>
<tr>
<td>Performance-Approach Goal Orientation</td>
<td>6, 9, 11, 12, 25</td>
</tr>
<tr>
<td>Performance-Avoid Goal Orientation</td>
<td>17, 24, 27, 28</td>
</tr>
<tr>
<td>Task Value</td>
<td>3, 8, 14, 19, 23</td>
</tr>
<tr>
<td>Control for Learning Beliefs</td>
<td>2, 7, 15, 21</td>
</tr>
<tr>
<td>Self Efficacy for Learning &amp; Performance</td>
<td>4, 5, 10, 16, 26</td>
</tr>
</tbody>
</table>