INTELLIGENT TUTOR SYSTEMS ADDRESSING STUDENT DISENGAGEMENT:
ADDING FORMATIVE REAPPRAISAL TO ENHANCE ENGAGEMENT AND LEARNING

By

Timothy J. Xeriland

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Educational Psychology and Educational Technology–Doctor of Philosophy

2018

ABSTRACT

INTELLIGENT TUTOR SYSTEMS ADDRESSING STUDENT DISENGAGEMENT: ADDING FORMATIVE REAPPRAISAL TO ENHANCE ENGAGEMENT AND LEARNING

By

Timothy J. Xeriland

This study examined the effectiveness of two different interventions designed to reduce student disengagement when using an Intelligent Tutoring System (ITS) to learn college-level mathematics. Both interventions used an emotion regulation strategy called formative reappraisal (FR), which used formative feedback processes to deliver just-in-time cognitive reappraisals to students about their performance. The first FR intervention used a utility value reappraisal, which was designed to help students reframe their understanding of the value of the current learning task. The second FR intervention used a cognitive reappraisal designed to help students better understand the difficulty and challenge of the current task.

The study was conducted with community college students (N = 136) randomly assigned to one of three groups (utility value FR, cognitive challenge FR, and no treatment condition) such that each group had students with a strong comprehension of geometry and those with a weak understanding. Three dependent variables were used to evaluate the effectiveness of the FR interventions: affective engagement, behavioral engagement, and learning.

Results show that students in both FR conditions had a statistical improvement in affective engagement and behavioral engagement in the first cycle of lessons compared to the no treatment group, but not in the second and third. Also, the post-test scores, used to measure learning outcomes, were significantly higher for the FR conditions compared to the no treatment condition. Additionally, no interaction effects were found between FR conditions and students with strong and weak geometry comprehension.

The study results reinforce the associations and mediating mechanisms proposed by Pekrun's control-value theory. Using the FR emotion regulation as an approach to influence students’ cognitive control and value appraisals was somewhat beneficial to these learners based on measurements of the dependent variables. Findings regarding FR suggested an emotion regulation strategy that uses a mix of antecedent-focused and response-focused strategies can decrease disengagement and increase meaningful learning for students using an ITS. There is a rapid growth in school systems using ITSs to deliver instruction, and these findings lay the groundwork for further studies to explore improving human-computer interaction.

To my grandfather, John Foley
who insisted no one takes you
seriously without an education.

ACKNOWLEDGEMENTS

To complete this odyssey required a Herculean effort, but I could not have achieved it without support. First, thanks to my dissertation committee for providing thoughtful direction during the process. I am especially grateful to Dr. Jack Smith and Dr. Lisa Linnenbrink-Garcia, who both spent many hours of patient guidance on steering me through the travails of this journey.

Also, thank you to faculty at Carnegie Mellon University, especially Dr. John Anderson and Dr. Vincent Aleven. Thanks to Dr. John Pane of the RAND Corporation and Dr. Sidney D'Mello of Columbia University.

Thanks to those who helped from Cedar Valley College. I am especially grateful to Tim Sonnier and Christa Crawford.

Many additional faculty members, staff members, and doctoral students have contributed to this project in ways both big and small. The entire 2010 EPET cohort showed incredible support. In particular, I want to thank Andrea Zellner who escorted me through the early years. Tracy Russo who has been the most faithful friend anyone could ask to have. April Joy Harrison, thank you for your love and support while leaning into the process with me. I will not take these things for granted.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

INTRODUCTION

In the spring of 2000, a small startup company approached the Dallas County Community College District (DCCCD) to introduce its product. The company had no clients, but the sales people figured if they could get a district as influential as DCCCD to use their software others would follow. To sweeten the deal, they offered their product free for life. DCCCD accepted, liked the results, and spread the word. That once small company is Blackboard Inc., which was instrumental in ushering in the online learning revolution.

In 2015, the DCCCD discovered an innovative technology to benefit learners called Intelligent Tutor Systems (ITSs). These computer tutors use artificial intelligence (AI) to model students' cognitions and prior knowledge to personalize instruction. The results from a trial run convinced college faculty and administrators to expand the use of ITSs to the mathematics program. In 2017, classes utilizing ITSs have seen an increase in both retention and academic performance compared to both previous years and traditional courses. As word of this success spreads to other institutions, ITSs are poised to explode in popularity.

A problem with these computer tutors is they are not effective for all students. In an ITS classroom, most of the interaction occurs between learner and computer. The teacher's role becomes a guide—they periodically review and discuss progress with students. With this setup, teachers discovered some students spend significant amounts of time logged into the ITS but made little advancement. When students become disengaged, the ITS is not sensitive to the issue the way a human tutor would react to pupils not paying attention.

My career in education started as a mathematics tutor and for a session to be effective it was vital to consider the learner's engagement level on an equal footing with their cognitive aspects, such as learning and memory. Researchers, practitioners, and policymakers alike have widely recognized meaningful learning requires sustained engagement (D'Mello, Dieterle, & Duckworth, 2017; Loveless, 2015; PISA, 2012). Intelligent tutor systems' inability to detect the crucial element of engagement is a concern because it is a precursor to the learning process, and research shows it is significant in effortful problem-solving and deep thinking (Csikszentmihalyi, 1990; D'Mello, Olney, Williams, & Hays, 2012; Lepper, Woolverton, Mumme, & Gurtner, 1993).

The broad focus of ITS research has been on improving the model of how students organize knowledge to produce intelligent behavior and then developing a computer system to effectively interact with the learners' cognitive architecture (VanLehn, 2011). The ITS field is rooted in mechanism, and it treats tutoring like an industrial process (Woolf, 2009). The belief is if only programmers can fine-tune the algorithms, the tutoring session will hum along perfectly. As an experienced tutor, I would argue what distinguishes human tutoring from other forms of instruction, is the highly personal nature of the activity. A productive session requires an active tutor to develop an integrative perspective of the student recognizing the whole person and not merely a disembodied mind in a classroom.

Recently, some researchers have built computer learning systems designed to detect disengagement (Afzal & Robinson, 2009; Burleson & Picard, 2007; D'Mello et al., 2012). Much of the work has focused on the mechanics of detecting student disengagement, while strategies to improve the learner's focus, after detection, are still in their infancy. For example, one study investigated promoting engagement and learning in an ITS by reacting to student gaze patterns. D'Mello and colleagues (2012) used a sophisticated commercial eye tracker to monitor students' eye-movements for gathering data to detect when students became disengaged. While the technology was state-of-the-art, the intervention included simple statements made by the ITS (e.g., “please pay attention”). Although this procedure did not influence student motivation or self-reported engagement, it was enough to produce some learning gains (D’Mello et al., 2012).

The current growth of ITSs is dramatic and has the potential to radically change the educational process as the technology spreads (Koedinger, Brunskill, Baker, McLaughlin, & Stamper, 2013). This year, IBM is launching an ITS using its advanced natural language processing system, known as Watson, in higher education and preschool both online and in classrooms (IBM Pearson, 2016). As schools integrate these systems, it is important to understand that while ITSs have yielded remarkable results in supporting learning (Evans & Michael, 2006; Reif & Scott, 1999; VanLehn et al., 2007), the systems do not address some issues relevant to student success as effectively as human tutors. To develop tutoring systems genuinely mimicking human tutors, ITSs would benefit from responding to learners’ apparent disengagement beyond a simple request to pay attention.

Statement of Problem

Building automated systems to detect disengagement is technologically challenging. Efforts to leap this hurdle have overshadowed research questioning the specific computer-based interventions the ITS should perform after the detection has occurred. The more challenging question is, once a system detects disengagement, what actions performed hold the most promise for reengaging students? In this study, I examine different disengagement-detecting interventions using theory that integrating cognitive and emotion factors to increase the potential for student reengagement.

Overview of the Dissertation

Chapter 2 presents a theoretical basis for the paper and examines existing research to develop an appropriate intervention. Chapter 3 details a pilot study conducted and how that informed the main study. Research questions that guide the main study are provided at the end of this chapter. Chapter 4 describes the research method of the study in detail. Chapter 5 provides the results where the three dependent variables are investigated with a series of ANOVAS and MANOVAS. Chapter 6 gives a general discussion, concentrating on the broader perspective of the study's findings, then addresses the implications and limitations of the study, as well as future directions.

CHAPTER 2:

THEORY AND EMPIRICAL RESEARCH

Theory and past empirical research provide an understanding of factors shaping the study. I first situate the reader into the broader context of the problem to be solved. Then, I break the chapter into three sections: a) theory, b) regulation, and c) disengagement. First, the theoretical basis is explicitly revealed to allow for a broader perspective in understanding the research questions and design. Second, using the existing research, I present an approach to address the problem. Finally, an investigation of past studies helps to consider reasons for modifying the intervention based on whether the student is high- or low-performing.

Human Tutoring and Computer-Based Tutor Systems

Individual tutoring is one of the oldest forms of instruction and is still one of the most effective methods by which students learn (Bloom, 1984). Researchers recognize having a human tutor is significantly more effective than mass instruction done through classroom teaching, yet it is not economical to provide every student with a tutor (Graesser, VanLehn, Rose, Jordan, & Harter, 2001). A promising solution emerged with the advent of computers, and programmers immediately sought to develop computer tutors that could mimic human counterparts (Smith & Sherwood, 1976).

Early efforts followed an approach where computer tutors gave students feedback directly on their answers (Skinner, 1958). For example, learners finding the length of a triangle would draw it on scratch paper, calculate the answer, enter the result, and receive either a congratulation or a hint (VanLehn, 2011). These first-generation computer tutor systems are commonly referred to by the retronym *Computer Assisted Instruction* (CAI). Researchers discovered CAI produced positive effects compared to no tutoring but fell considerably short when likened to human tutors (Kulik & Kulik, 1991a).

As *artificial intelligence* (AI) emerged, computer tutors evolved past traditional CAI. The transition from CAI toward *intelligent tutor systems* (ITSs) arrived when these new tutors could pass three tests of intelligence (Burns & Capps, 1988). The three types of knowledge and problem-solving expertise are a) expert knowledge of the subject—ability to solve problems in the domain, b) student diagnostic knowledge—ability to deduce an approximation of the learner's understanding, and c) instructional knowledge—ability to implement strategies to improve student learning (Burns & Capps, 1988). These new tutor systems, grounded in AI and cognitive theory, carefully guide learners through problems by providing feedback drawing on their three types of intelligence. Researchers developed CAI to provide an evaluation of student solutions. The new intelligence built into ITSs, however, began to shrink the granularity of feedback down to individual steps taken to reach answers, improving the effect size of these tutor systems (Evans & Michael, 2006; Reif & Scott, 1999; VanLehn et al., 2007).

Intelligent tutor systems have made remarkable progress in a relatively brief period (Crowe, LaPierre, & Kebritchi, 2017). The latest meta-analysis by Kulik and Fletcher (2016) investigated 50 empirical studies and concluded ITSs raised student test scores above both CAI and human tutors. ITSs appeared to be particularly adept at improving student mathematics performance (Beal, Arroyo, Cohen, & Woolf, 2010; Beal, Walles, Arroyo, & Woolf, 2007). Researchers, however, have not reached an agreement on the effect size for student learning. Steenbergen-Hu and Cooper (2013) investigated mathematics in K-12 and revealed no overall effect on learning in another meta-analysis review. Researchers conducted a large-scale study and discovered many students receiving algebra training from an ITS were still unable to pass the course from ITSs training alone at the end of a 24-week semester (Campuzano et al., 2009). Kulik and Fletcher (2016) sum up this controversy with the statement, “The lack of consensus about ITSs effectiveness is striking” (p. 46).

A study done at Cedar Valley College could shed light on the conflicting results found in ITS research. As part of the college’s Quality Enhancement Plan, staff investigated mathematics students using the ITS known as ALEKS (for Assessment and LEarning in Knowledge Spaces). Administrators and faculty divided students into traditional courses and ALEKS courses. At the end of the first year, they realized students in the classes that used ALEKS had better averages on measurements such as test scores, completion rates, and success rates (SACS, 2018). One example is a developmental mathematics course called DMAT 0310. After years of experimenting to increase the completion rate in DMAT 0310, the faculty were able to improve it to 90.4% in a traditional classroom. After only one-year of using ALEKS, the same course increased to a 95.5% completion rate (SACS, 2018). Unfortunately, an examination of other mathematics courses was not as promising. The class DMAT 0305 had a student success rate of 46%, in traditional classrooms, and while using ALEKS did show some improvement, it was still less than ideal success at 53.5%. The question became why were nearly half the student in the ITS course unsuccessful? The generated reports from ALEKS revealed that students who were unsuccessful spent a vast amount of time in the system but did not complete modules or make progress. The ALEKS summaries indicate the ITS worked well, except for learners who were unengaged.

To improve ITSs, researchers prioritized developing systems that provide students with more detailed feedback. Scholars refer to this approach as the *interaction granularity hypothesis*, and it presumes ITSs interacting with increasingly smaller pieces of student cognition will produce more useful feedback and explanations (VanLehn, 2011). This approach is problematic. Studies find increasing the granularity of feedback (moving ITS input on answers from steps down to substeps) does not improve results and could decrease the effect size for the systems (Evans & Michael, 2006; Reif & Scott, 1999; VanLehn, 2011; VanLehn et al., 2007).

Most developments with ITSs have been one-sided, focusing on student learning and cognition, while ignoring emotions and engagement (D'Mello et al., 2012). The uneven concentration is a serious limitation as any student can testify how their emotions influence engagement in learning the material. Anxiety before a test, boredom during a lecture, and frustration at failure can influence a student's attention, engagement, and study strategies used to learn new information and skills (Artino, Holmboe, & Durning, 2012).

Theoretical Background

The study integrates across multiple theories to develop a system to engage students. Specifically, I draw from the *control-value theory* of achievement emotions (Pekrun, 2006) and a *schema* framework of learning (Piaget, 1976). Integrating core ideas from these two theories is key to understanding the treatment chosen to address student disengagement with intelligent tutor systems.

Control-Value Theory

Historically, educational psychologists have devoted few resources to understanding the role student affect has on learning (James, 1890; Pekrun, & Linnenbrink-Garcia, 2014). Theorists have ignored affective states such as boredom, frustration, and disengagement in most classic models of cognition (Dweck, Mangels, & Good, 2004). Indeed, researchers began to theorize that academic learning was chiefly a cognitive activity free from the burden of emotions (Brown, Bransford, Ferrara, & Campione, 1983). Recently, scholars concerned with the lack of unification between cognition and emotion have requested a more comprehensive approach to human learning (Linnenbrink & Pintrich, 2004). There is a definite need to integrate cognition and emotion to understand how a student arrives at academic success. To address the gap, I have turned to a unified model, known as the control-value theory of achievement emotions, which integrates cognition and emotion to understand how a student arrives at academic success (Pekrun, 2000; Pekrun, 2006; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010; Pekrun, Goetz, Titz, & Perry, 2002).

A fundamental concern for educational psychologists such as Pekrun (2000) involves the processes determining the many potential performance outcomes for students. Pekrun's (2006) control-value theory posits once teachers give students a task, they will consider how important it is to them (value) and whether they have the skills required to complete it (control). The method of students mentally weighing the prospects of a task is referred to as a *cognitive appraisal* (Pekrun, 2006).

When a student makes a cognitive appraisal of an assignment, whether conscious or unconscious, deliberate or involuntary, it serves as a catalyst for determining the emotions that arise (Gross, 2008; Johnson-Laird & Mancini, 2006; Pekrun, 2006; Schutz & Davis, 2010). Pekrun (2006) refers to these as *achievement emotion*s, which follow from engaging in achievement activities (e.g., studying mathematics) or their subsequent outcomes (e.g., grade on a test).

In control-value theory, two types of appraisals influence achievement emotion during learning: *perceived control* and *perceived value* (Pekrun, 2006, 2009). Perceived control is the degree to which students believe their effort can alter the outcome (Pekrun, 2006). For instance, if a student thinks no matter how hard he studies for a mathematics test he will still fail, then his perceived control is low and negative achievement emotions such as frustration are more prevalent (Pekrun, 2017).

Perceived value is the importance a student places on an activity or outcome. If the perceived value is low, negative emotions can result as the student cares little for the material (Pekrun et al., 2010). For example, if a student listening to a lecture on Euclid's fifth postulate fails to see the relevance to their lives (low value), emotions such as boredom are likely to follow (Pekrun, 2006).

Value and control appraisals do not operate in isolation but interact with each other (Pekrun, 2006). Take an example of two students with low control appraisals for learning mathematics. The first student, with little regard for the topic, may experience the achievement emotions of apathy or boredom on the task, while the second student, who greatly values the subject, may experience anger or frustration at failing.

Pekrun's (2006) guiding theoretical framework is captured in a diagram (Figure 1). The model lays out a chain of events involving cognitive, emotional, and behavioral responses resulting in academic achievement. The complete process resides inside a learning environment where distal factors can influence elements of the process (Pekrun, 2006, 2009). These distal factors can include the speed the tutor presents the material, the emotional support the tutor offers, or the number of students in a session. For instance, a student working on an intelligent tutor system alone in a lab can expect a different academic experience than if she studied the material in a classroom full of students.

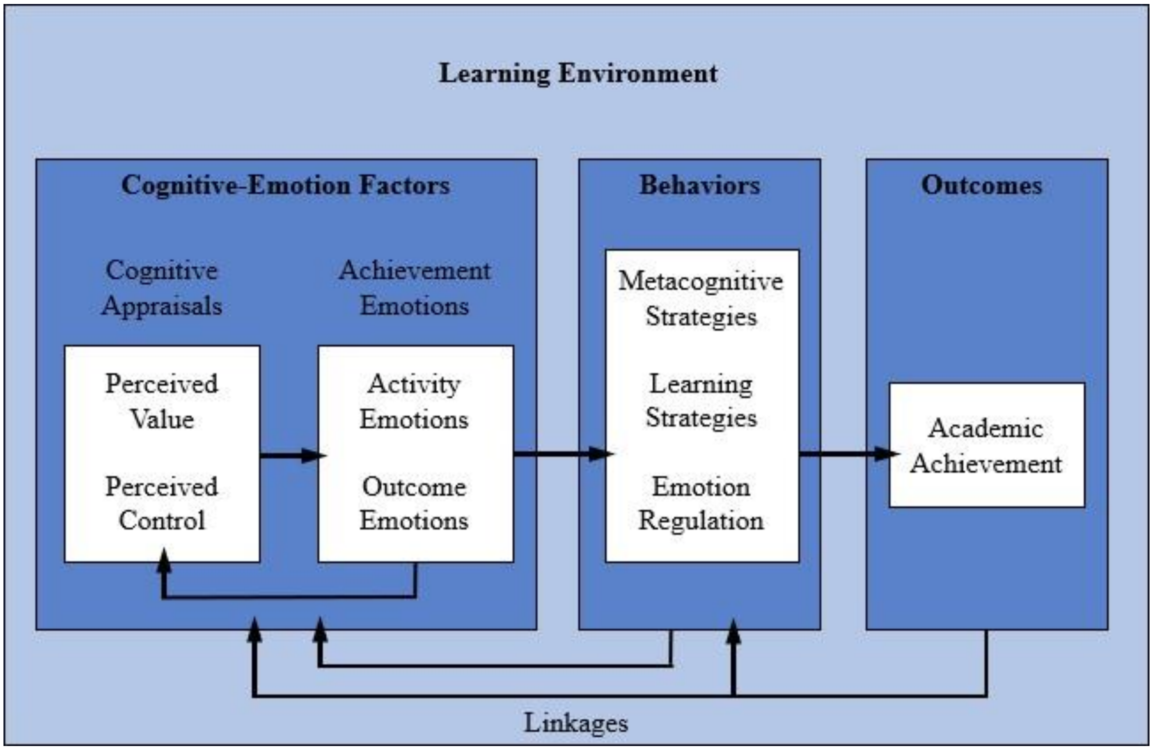


Figure 1. The guiding theoretical framework (adapted from Artino et al., 2012; Pekrun, 2006).

The model in Figure 1 above appears to run in one direction (starting with cognitive appraisals down to academic achievement), but the process is bidirectional. The arrows labeled as *linkages* show the interdependence of the components. For example, appraisal directly influences emotions, but the emotions generated are fed back to influence any further weighing of the task. The same relationship applies to the rest of the model. In sum, the learning environment, appraisals, emotions, strategies, and outcomes are linked to the student learning process and occur reciprocally (Pekrun, Frenzel, Goetz, & Perry, 2007).

To better understand Pekrun's (2006) model, it is insightful to apply it to traditional educational approaches such as ITS development. The diagram (Figure 1) shows a layer, called *behaviors*, between achievement emotions and the resulting outcome. Intelligent tutor system development has concentrated on directly providing feedback on *learning strategies* contained in this layer (VanLehn, 2011). The model suggests and is backed by research that even if investigators develop useful learning strategies to provide the students, the specific emotions experienced can influence whether the learners use them (Aspinwall, 1998). Students who become bored, confused, or frustrated, may simply shut down and not use any approach previously taught. This situation is why merely teaching students learning strategies is problematic if the appropriate antecedent achievement emotion or motivation is not present to compel the student to use the technique (Pekrun et al., 2007). Likewise, changing an achievement emotion is complicated without attention to student appraisals (perceived control and perceived value), because of their influence on emotion (Pekrun, 2006). By switching the focus to more fully understand the role achievement emotions and their proximal determinants, appraisals, we can see why a tutor (human or computer) needs to consider other aspects besides only learning.

Control-value theory provides a system for guiding predictions on how cognitive appraisals, achievement emotions, and behaviors, influence academic achievement. But, to make lasting improvements, steps need to be taken to decide how to intervene and stop or improve harmful appraisals and emotions. The approach I reveal depends on a framework for learning known as schemas, which I cover next.

Schema Framework

Some educational psychologists postulate that learners make sense of new content by incorporating the information into their existing mental *schemas* (Howes & O’Shea, 2014). These schemas serve as a framework for representing how we organize our thoughts around different concepts and how we interpret the latest information. Philosopher Immanuel Kant proposed the initial idea of schemas as innate structures used to help us perceive the world (Nevid, 2007) and the concept was later expanded on by Piaget (1976).

In Piaget’s (1976) view, a schema is developed or modified through two different processes: assimilation or accommodation. Assimilation is the process of adding additional information to an existing schema (Piaget, & Cook, 1952; Siegler, DeLoache, & Eisenberg, 2003). Once learners form schemas, they need to see many examples to build up their *schematic network*. For example, once a child learns the concept of a triangle, he will need to see many cases of assorted styles to assimilate into his triangle schema and to properly learn the concept (Piaget, & Cook, 1952; Siegler et al., 2003). As researchers have documented, students not exposed to enough examples of geometric figures can fail to classify those shapes later (Burger & Shaughnessy, 1986).

Accommodation involves the learner developing an understanding of novel concepts through altering existing schemas or creating entirely new schema to house the information (Auger, & Rich, 2007). For example, the child who learned about triangles may confuse the form with squares, and proper instruction develops a second schematic framework for the new concept thus allowing the child to accommodate the new information.

This building process does not mean the student will come to an understanding of geometry postulates, theorems, or proofs, wholesale à la Plato's Meno, but through a series of discrete acts of learning. Taber (2010) suggests a schematic network may restrict learning in three ways:

- New information can only be considered by the student in a limited amount;
- New information must in some way connect to prior knowledge;
- New information only becomes permanent hours after the initial presentation because it requires substantial restructuring of current knowledge.

The concepts of a schema framework and control-value theory are key to understanding the theoretical underpinnings of the paper. The following section discusses how combining these theories leads to an intervention to modify student appraisals.

Integrating Across Theories

In this section, I join control-value theory with the concept of schemas, which helps identify a promising technique to influence student appraisals. In control-value theory, positive student appraisals are important to academic success (Pekrun, 2006). Reisenzein (2001) suggests a key to influencing these appraisals is to treat them as a type of schema. This understanding leads to an important question: Can changing a student's appraisal schema, through either accommodation or assimilation, improve achievement emotions, which in turn improve academic performance?

Researchers consider changing schemas to involve simultaneously weakening old schemas and strengthening new ones (Padesky, 1994). Gross (2008) proposes methods that seek to alter students' existing appraisals by guiding learners to generate new meanings in interpreting their judgments (Gross, 2008). This restructuring process is precisely the method suggested to adjust schemas. From the control-value theory perspective, if it is possible to change how a student interprets a task appraisal, it is possible to change the inevitable emotions arising from the evaluation (Pekrun, 2006).

To tie these two frameworks together, I leveraged Gross' (2008) strategy known as *cognitive change*. According to Gross (2014), cognitive change refers to "modifying how one appraises a situation so as to alter its emotional significance, either by changing how one thinks about the situation or about one's capacity to manage the demands it poses" (p. 10). This system provides a feasible approach to change students' appraisal schemas of value or control while using an ITS. According to the control-value theory, improving appraisals brings an emotional improvement that, in turn, improves student academic achievement. The following section leaves the theoretical and I begin to unpack specific methods for students to regulate their appraisal schemas.

Self-Regulation

This section addresses self-regulation in students. The focus of self-regulation outlined here is that these behaviors develop through a process of students creating mental schemas through habitual practice that empowers students to automate academic skills (Winne, & Jamieson-Noel, 2003). Intelligent tutor developers have begun to investigate how self-regulation influence students' work in the ITS environment (Azevedo, Cromley, & Seibert, 2004a). I divide this section into three parts: self-regulation learning, emotion regulation, and formative reappraisal.

Self-Regulated Learning

Researchers have demonstrated a link between providing students with instructional aids encouraging them to self-regulate their cognitive and affective states, and improved performance when studying in computer-based environments, including ITSs (Azevedo et al., 2004a). *Self-regulated learning* (SRL) is the use of strategies that allow the student to take control of and evaluate their thinking, affect, and behavior (Pintrich, 1995). One SRL strategy is metacognition or thinking about one's thinking. Flavell (1976) provides an example, "I am engaging in metacognition if I notice that I am having more trouble learning A than B; if it strikes me that I should double check C before accepting it as fact" (p. 232). Research shows students rarely use SRL during learning in a computer-based environment (Azevedo, Guthrie, & Seibert, 2004b; Greene & Land, 2000).

One study demonstrated the effectiveness of university students using SRL in a computer-based environment. Azevedo and Cromley (2004) randomly assigned 131 undergraduate students to either an SRL group or a control group. The students learned about the circulatory system using the tools in a computer-based environment. The result empirically demonstrated that students who received the SRL training improved both their performance and ability to process data. Also, the SRL group gained a deeper understanding of the material compared to the control group. Both groups of students learned declarative knowledge in the study, but for higher-order questions, which requires learners to have an accurate mental model of the circulatory system, only the SRL group showed significant improvement (Azevedo & Cromley, 2004).

Students are required to engage in SRL when transitioning from a highly structured, teacher-led classroom to the individualized learning environment found in ITSs (Winters, Greene, & Costich, 2008). Without the presence of a teacher, students are solely responsible for monitoring their cognitive, emotional, and behavioral condition. For students' cognitive concerns, the ITS platform typically facilitates the learning process by providing ample scaffolding (Belland, Walker, & Kim, 2017). The concern is instruction necessitates more than interacting with student cognitions, and human tutors will attest that coaching requires attending to student behavioral and emotional issues that can interfere with learning. I cover a method of SRL that applies to emotions in the following section.

Emotion Regulation Using Reappraisal

Teaching emotion regulation strategies to students has been demonstrated to aid in student learning (Dillon, Ritchey, Johnson, & LaBar, 2007). The collective experience of politely smiling upon receiving an unwanted gift is an example of emotion regulation, which Gross

(1998) defined as “the process by which individuals influence which emotions they have, when they have them, and how they experience or express these emotions” (p. 275). The emotion regulation mechanism can be as simple as cheering yourself up by thinking about a happy moment or controlling your breathing when stressed. Individuals who use regulation have control over whether an emotional reaction is desirable or should be deferred (Cole, Michel, & Teti, 1994). Ninety percent of undergraduates’ report using techniques to modify their emotions daily (Gross, 1998; Gross, Feldman Barrett, & Richards, 1998). Although, researchers have found students in computer-based environments have difficulty regulating their emotions on their own (Winters et al., 2008). The two most investigated and agreed upon emotion regulation strategies are *cognitive reappraisal* and *expressive suppression* (Cutuli, 2014; Gross & John, 1998).

Cognitive reappraisal is an *antecedent-focused* strategy used before complete activation of the emotional response has occurred (Cutuli, 2014). The process includes two parts: a) recognizing an emotional situation will likely elicit a reaction (appraisal), and b) reframing the situation (reappraisal) to move in a neutral or positive direction (Gross & John, 2003; Lazarus & Alfert, 1964). Those who are adept in this process can take an unpleasant situation and completely upend any negative experience by quickly recasting it to a more agreeable circumstance (Sun, 2014). What was once a reckless driver, is now only someone trying to get to work on time.

Expressive suppression is a *response-focused* strategy that intervenes after an experience activates an emotion and the behavioral response is fully produced (Cutuli, 2014). The process is an attempt to mask, inhibit, or reduce the emotion-expressive behavior we show to others through facial expressions or conduct (Gross & John, 2003). A person may experience the emotion but avoid responding. For example, after getting cut off in traffic, a person may encounter a flash of anger but consciously avoid reacting behaviorally to the emotion. Some researchers discover that expressive suppression can impair cognitive abilities (Sheppes & Meiran, 2007, 2008). One study confirmed that students who suppressed their emotions during a lecture, compared to students who did not, later performed worse on the exam (Richards & Gross, 2000). Additionally, expressive suppression negatively influences memory performance (Richards & Gross, 1999; Hayes et al., 2010). Any intervention that diminishes cognitive ability is not a suitable candidate for an academic environment.

Cognitive reappraisal improves affect while not impairing cognitive or physiological conditions (Dillon et al., 2007; Ehring, Tuschen-Caffier, Schnülle, Fischer, & Gross, 2010; Germain, & Kangas, 2015; Richards & Gross, 1999; Shiota & Levenson, 2009). Studies show that cognitive reappraisal has enhanced students' memory performance (Dillon et al., 2007; Richards & Gross, 1999). It has been suggested that reappraisal promotes deeper cognitive analysis by producing more elaborate memory traces known as the levels-of-processing effect (Craik & Lockhart, 1972; Dillon et al., 2007). This finding reinforces the advantage of using reappraisal as an emotion regulator in an educational setting. Since most students do not possess the tools to use this strategy independently, investigators often directly instruct students to perform cognitive reappraisal. Gross (2014) refers to this as *instructed reappraisal*. For example, a teacher can request a student use his or her imagination to view a performance task from someone else's perspective with the aim of depersonalizing the task and reducing the negative affect.

An illustration of instructed reappraisal is when participants watched a nauseating medical training film. Goldin, McRae, Ramel, and Gross (2008) taught one group to imagine they were medical professionals viewing the film. The group using the reappraisal strategy experienced less negative affect than the group with no instruction. Another study had depressed patients watch a sad movie (Ehring et al., 2010). One group used instructed reappraisal while the other used expressive suppression. Students in the reappraisal group imagined being the director and viewed with an eye for the technical aspects of the film such as using cuts and camera angles (Ehring et al., 2010). When the investigators compared the two groups, those in the emotion suppression group had a significantly higher heart rate during the film and considerably higher levels of persisting negative mood after the movie compared to the instructed reappraisal group (Ehring et al., 2010). The research studies suggest instructed reappraisal is a potentially helpful intervention when dealing with negative affect. Surprisingly, researchers have done little to adapt the technique for improving student affect or engagement in educational settings (Strain & D'Mello, 2015).

One plausible reason for instructed reappraisal's conspicuous absence from the classroom could be that Gross (2008) developed it as an antecedent intervention. The antecedent approach means experimenters apply it before any negative behavioral consequences such as disengagement occurs. While the preemptive treatment may help with the initial appraisal of an academic task, most assignments are complex and involve many subtasks for students to complete. In other words, a learning experience is rife with students making appraisals (Bieg, Goetz, & Hubbard, 2013). As it is essential for instructors to provide feedback on errors in student thinking, it is equally important to deliver feedback on emotional and behavioral issues. The emotion regulation strategy cognitive reappraisal holds promise for improving negative behavioral states like disengagement. Developing a system regularly encouraging students to reevaluate their appraisals could prove beneficial. The next section includes the research of best practices in supplying feedback to students about their appraisals.

Regulation: Formative Reappraisal

Research reveals feedback is one of the most powerful influences on learning (Hattie & Timperley, 2007; Shute, 2008). Most inquiry on feedback regards its effects on learning while work on feedback for appraisals is scant. A review by Daniels and Gierl (2017) found only five studies investigating feedback for achievement emotions, and none of these examined how feedback influences students' appraisals (Fong et al., 2018). Typically, feedback is given to improve a student's cognitive (problem-solving) approach to course content or to provide learning strategies. By also supplying feedback encouraging reappraisals, it can benefit students through modifying their affectual and behavioral responses (e.g., stopping disengagement from occurring).

Instructed reappraisal is an antecedent intervention, which means researchers do not reapply in response to a behavior (Gross, 2008). While ITSs have used instructed reappraisal with moderate success (Strain & D'Mello, 2015), if a tutoring system is reactive to student disengagement it can deliver a booster shot of the intervention by providing students with feedback to engage in additional reappraisals as needed. This just-in-time reappraisal (called *formative reappraisal* (FR) from here on) is a type of delivery method required for a reactive ITS. Using FR can begin with an antecedent-focused strategy, although adding *formative feedback* also makes it a response-focused strategy.

Formative feedback is defined as "information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning" (Shute, 2008, p. 154). An example of formative feedback is a teacher observing and helping students while they work on problems and then redirecting students if the teacher notices errors in their thinking. The literature on providing student feedback strongly supports a formative approach, which has been demonstrated to improve student performance (Black & Wiliam, 1998), motivation (Narciss & Huth, 2004), and affect (Schwartz & White, 2000). Research suggests formative feedback has advantages for complex activities found in learning environments (Popham, 2011) and these intricate learning settings mean students are making many appraisals requiring feedback.

Investigators have studied the best methods for delivering formative feedback. Effective formative feedback is descriptive, supportive, well-timed, specific, credible, confidential, sensitive, and infrequent (e.g., Brophy, 1981; Schwartz & White, 2000). ITS investigators have worked for decades experimenting with differential feedback timing effects on student learning (Anderson, Corbett, Koedinger, & Pelletier, 1995; Corbett & Anderson, 2001). Some studies show the most significant learning gains while using an ITS occurred for students who consistently received immediate feedback (Corbett & Anderson, 2001; Dihoff, Brosvic, Epstein, & Cook, 2003; Phye & Andre, 1989). Also, Schofield (1995) witnessed immediate feedback improves student motivation. At the same time, overwhelming students with feedback that is too frequent is detrimental (Brophy, 1981; Schwartz & White, 2000). The research suggests immediate feedback on disengagement is best, but instructors should apply sparingly; otherwise, feedback will not allow the students to sort through any of the processes on their own (Wiliam, 1999).

Recent studies bolster the FR strategy by demonstrating the benefits of providing ITS students with immediate feedback that encourages the use of regulation strategies (Feyzi-Behnagh et al., 2014; Trevors, Duffy, & Azevedo, 2014). El Saadawi and colleagues (2010) investigated the performance of medical students using the SlideTutor ITS. The researchers provided either immediate feedback or feedback that was faded (reduced as time passed) on regulation strategies (i.e., metacognitive scaffolds). Results showed that participants receiving immediate feedback significantly improved metacognitive performance as well as learning gains. Fading of immediate feedback had the opposite effect and led to a decreased metacognitive performance regarding accuracy and discrimination (El Saadawi et al., 2010).

The reason for instructing students to use a reappraisal strategy is that many have not developed the skills to use it on their own (Ehring et al., 2010). In a similar vein, the reason to utilize a feedback system requesting reappraisals throughout the process is that many students have not yet developed a habit of using the emotion regulation when needed (De la Fuente & Cardelle-Elawar, 2009). Instructors helping students cultivate this practice is ideal because the habitual use of reappraisal improves cognitive performance and memory (Ortner, Ste Marie, & Corno, 2016). Formative feedback is believed to develop students to become self-regulated learners because the strategies it uses naturally enhance the regulation process (Nicol & Macfarlane-Dick, 2006; Sadler, 1989). Additionally, research shows formative feedback helps students develop habits through a process of timely reminders (Nahadi, Firman, & Farina, 2015). Instructors who provide feedback encouraging learners to engage in reappraisal can assist them to develop an emotion-regulation habit—the goal for FR.

The stepwise process of many academic tasks suggests a reappraisal intervention may need to occur multiple times throughout a session (Bieg et al., 2013; McRae, Ciesielski, & Gross, 2012). FR is promising because it provides a possible approach to systemically upend negative appraisals adversely influencing achievement emotions and the resultant student behavior. While planning to intervene to assist students, it is important to note that evidence suggests the achievement emotions of high- and low-performing students may be different and therefore may require different reappraisals (Blayney, Kalyuga, & Sweller, 2010; Sweller, 2005). The next section explores these differences.

Disengagement

Students who become disengaged and are unable to sustain focus struggle to make meaningful learning gains (Loveless, 2015; PISA, 2012). Students' disengagement from activity in ITSs is seen as particularly problematic and requiring more study (D'Mello et al., 2012). I break this section into three parts, and each considers these questions about disengagement: a) Do certain types of students get disengaged for specific reasons? b) Is there a preferred approach to assisting low-performing students who get disengaged during ITS training? And c) Is there a preferred approach to helping high-performing students maintain focus during ITS practice?

Student Variations in Disengagements

Educational psychologists have long recognized the need to consider the cognitive differences between high- and low-performing students (Kulik & Kulik, 1984, 1985, 1991b, 1992; Rogers, 1991) and this section I make the case to consider emotional differences in these learners as well. Researchers who use the control-value theory suggest student achievement emotions can produce behavioral responses, like disengagement (Pekrun, 2006). As a classroom teacher, I regularly witnessed how students' affectual experiences influenced their engagement levels. If emotions produce disengagement, the question is, do certain types of students experience similar affectual states?

A study of 389 university students realized strong correlations between a student's grade point average (GPA) and the amount of hopelessness and shame they tend to experience (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). Low-performing students were more likely to experience negative achievement emotions compared to their high-performing peers. Meanwhile, researchers discovered boredom was experienced more regularly by high-performers than other negative affective states (Pekrun et al., 2011). In general, broad measures of negative emotion are related to lower student GPA and achievement scores (Gumora & Arsenio, 2002); positive emotions such as joy, hope, and pride are correlated with students' overall achievement (Pekrun, Goetz, Perry, Kramer, & Hochstadt, 2004). If it is true activity-related achievement emotions are different depending on student skill level, would that difference matter in addressing disengagement?

Consider the idea of assisting students whose interest has waned by allowing them to briefly change their learning environment and engage in an unrelated task. Initially, researchers believed the off-task behavior resulted in diminished learning (Carroll, 1963). Later, investigators discovered halting an activity and temporarily altering the environment can re-engage students (Baker, 2007). Recently, the discrepancy was revealed to be likely caused by the affective state prior to the off-task behavior (Sabourin, Rowe, Mott, & Lester, 2011). The investigators discovered boredom before an off-task behavior leads to increased engagement when students return to the original assignment, whereas confusion before an off-task behavior leads to decreased engagement when moving back to the initial task (Sabourin et al., 2011). Additionally, research establishes low-performers are more likely to experience confusion (Gumora & Arsenio, 2002), and high-performers are more apt to experience boredom (Pekrun et al., 2011). Therefore, the skill level of the student would be a significant consideration when deciding on the intervention that pulls the student off-task—exactly the way FR does.

The *expertise reversal effect* adds substantial evidence why a learner's skill level can produce completely different outcomes for an intervention and why educational researchers need to consider the type of student in studies. Expertise reversal effect is when the usefulness of an instructional technique for novices loses its advantage and may *reverse* as levels of expertise increase (Kalyuga, Rikers, & Pass, 2012). The redundancy principle is one of the chief causes for reversals in the effectiveness of instructional techniques and shows the counterintuitive idea that adding supplementary material to instruction can reduce learning performance (Kalyuga & Renkl, 2010). For example, written explanations explaining fundamental mathematics concepts alone can produce better learning outcomes than when providing additional multimedia to accompany the printed material (Rey & Buchwald, 2011). Although, research shows learner variation for these tasks can make a difference with the principle. Novices improved learning performance when provided additional material compared to peers without the supplement; while the extra content offered, reduced performance for experts (Blayney et al., 2010; Sweller, 2005).

Some of the oldest and most well-established research in psychology shows having other people present during a task significantly improves performance (Triplett, 1898). The influence of outside observers to boost an individual's performance is known as *social facilitation*. In 1965, Zajonc discovered the presence of others did not always show task enhancement. Investigators realized benefits to performance occurred only when the person was skillful at the task (Zajonc, 1965). Research shows that low-performers of a given task usually do not experience social facilitation, and the presence of others can undermine them (Rosenbloom, Shahar, Perlman, Estreich, & Kirzner, 2007). Highly skilled Olympians set personal records in front of a cheering crowd while novices can find their performances wither under watchful eyes. The social facilitation phenomenon reveals the difference between high- and low-performers may not be the degree but the direction (i.e., what is helpful to one, is harmful to the other).

Researchers should consider the student skill level when addressing performance drops or disengagements with educational interventions. The next section covers what is believed to be the underlying cause of why low-performing students become disengaged when using an ITS.

Low-Performing Students' Disengagement: Task Value with ITSs

Cognitive load theory (CLT) proposes students have finite resources they can draw from during the learning process (Chandler & Sweller, 1992; Sweller, 1988; Sweller, 1999). According to CLT, intrinsic and extrinsic factors play a part in determining the amount of cognitive load a learner experiences (Sweller, 1988). Extrinsic factors come from the outside and how the instructor presents material can influence its load. For instance, if the instructions for changing a car tire are only given by reading a manual, it may increase the extrinsic load for students having to imagine the process. The intrinsic component weighs how difficult the material is along with a learner's ability to process information. For example, understanding integration taught in Calculus may be difficult for a student to learn, especially if he does not have the background. CLT proposes as the complexity of the material the student studies outstrips their ability to process the information it induces cognitive overload (Sweller, 1999). Several studies document the decreases in learning performance because of the cognitive overload (Chandler & Sweller, 1992; Jones & Macken, 1993; Blayney et al., 2010).

Fortunately for students who find mathematics challenging, ITSs are constructed to avoid putting students in cognitive overload (Najar, Mitrovic, & McLaren, 2016; Woolf, 2009). First, the AI has built-in algorithms (e.g., Genetic, ANN, and TF-IDF) that monitor student progress and create individual learning paths that purposely direct students to comprehensible material (Al-Radaei, & Mishra, 2011). Only a student demonstrating mastery of the topic allows the program to progress to novel content (Guzmán, Conejo, & Pérez-de-la-Cruz, 2007; Jeremić, Jovanović, & Gašević, 2012). Second, the ITS breaks problems down into steps where it continually evaluates student responses and provides scaffolding when needed, which is an advantage for learners experiencing high intrinsic cognitive load (Belland, Walker, Kim, & Lefler, 2017; Foshee, Elliott, & Atkinson, 2016; Lin, Guot, & Lin, 2016; Yarnall, Means, & Wetzel, 2016). The ITS environment offers extensive support options. For instance, a standard feature of an ITS is to provide hints and cues stepping students through to the solution, which is beneficial to learners who would typically find the material intrinsically challenging (Koedinger & Aleven, 2007). Researchers find ITS scaffolding is significant enough to produce the expertise reversal effect for high-performers, (Koedinger & Aleven, 2007; Salden, Aleven, Renkl, & Schwonke, 2008), which I cover in the next section.

Intelligent tutor systems provide the appropriate scaffolding and support options by removing intrinsic challenges novices typically face. Therefore, students negatively appraising the value of the topic would be the chief culprit of any damaging achievement emotions (Pekrun, 2006; Pekrun et al., 2002). Developers who program ITSs need to consider the level of challenge the material poses to students. Once the ITS delivers material adequately challenging any continued disengagement must reside in the value the student places in learning the content. Eccles et al. (1983) hypothesized there are four main elements students use to find value: a) utility value, b) perceived cost, c) incentive value, and d) attainment value. *Utility value* is relevant to this study. Utility value is the students' evaluation of how useful engaging in a task is for helping to meet their goals. For example, if a student is planning to drive to California in the summer, then the value of learning the material to pass a driver's test will likely be high.

Manipulating utility value has a positive effect on engagement and learning for low-performing students that is of primary interest. Researchers revealed increasing utility value in students significantly improves performance and engagement in educational environments (Cole, Bergin, & Whitaker, 2008; Durik, Vida, & Eccles, 2006; Hidi & Renninger, 2006; Hulleman, Durik, Schweigert, & Harackiewicz, 2010a; Hulleman, Godes, Hendricks, & Harackiewicz, 2010b; Simons, Vansteenkiste, Lens, & Lacante, 2004). Increasing utility value may not apply equally to students. Evidence suggests utility value can have different effects for students with different academic standings (Menager-Beeley, 2001). Students with *low prior grades* in English (i.e., low-performing students) perceived low task value compared to students with *high prior grades* (Menager-Beeley, 2001). Furthermore, increasing the task value for the low-performing group significantly reduced their dropout rate, while the researcher did not find this improvement for students with high prior grades (Menager-Beeley, 2001).

Utility value research done at community colleges in developmental classes (courses reserved for low-performing students) demonstrates the benefits for underachieving students. Parlett (2012) studied students in a reading course and discovered a strong influence of utility value on academic achievement ($r = .442$). Another study used a utility value intervention with mathematics students and realized learners who received the manipulation reported the topic was of more interest and value while improving learning performance compared to the control group (Gaspard et al., 2015).

The literature also suggests caution when manipulating utility value in low-performing students. Consider a study that investigated students with low perceptions of competence in mathematics compared to students with high perceptions (Godes, Hulleman, & Harackiewicz, 2007). The researchers discovered emphasizing utility value lowered the interest level for students with low competence perceptions while increasing interest for the student with high perceptions of competence. Additionally, Durick and Harackiewicz (2007) realized the same effect when using utility value on students with a low or high initial interest in mathematics. These studies show informing students how the content or task has personal value may not increase interest and it could lessen it for low-performing students. Hulleman and colleagues (2010b) proposed the intervention in these studies suffered from a flaw of researchers emphasizing their version of utility value, which may undermine low-performing students. These investigators suggest promoting utility value of mathematics by informing students how the material applied to their everyday lives is problematic because, “[f]or a student who does not do well in math or find math interesting, being told that math is important to his future may be threatening and intensify negative reactions” (Hulleman et al., 2010b p. 882). Utility value interventions done poorly can decrease engagement by causing students to further withdraw from the material.

Researchers suggest a more effective approach to promote utility value for low-performing students is by encouraging them to self-generate what they believe is the value of the learning content (Hulleman et al., 2010b). A self-generating approach gives students an opportunity to realize the value of the topic and find the highest relevance to their own lives. Hulleman and colleagues wrote, “Allowing students to discover the connections between the activity and their lives on their own may be a less threatening way to promote the perception of utility value, and it may therefore be particularly beneficial for students with low performance expectations” (p. 882).

For low-performing students, a nonthreatening way to manipulate utility value is with cognitive reappraisal that allows students to self-generate. A study of interest created an intervention combining elements of utility value with cognitive reappraisal (Strain & D’Mello, 2015). The researchers’ reappraisal approach had students consider the perspective of an outside observer as they learned about the United States Constitution and the Bill of Rights. The participants assumed the role of a person looking for a job and the experimenters asked them to consider the assignment as part of a job interview to increase the utility value. The investigators hypothesized having the students imagine a context where the task was personally valuable (finding a new job) would allow them to self-generate benefits without feeling threatened, considered a concern for low-performing students. Participants who used this approach reported more engagement and achieved higher learning outcomes (Strain & D'Mello, 2015).

While the Strain and D'Mello (2015) study discovered positive results, using reappraisal to imagine the stress of a job interview is likely problematic for some low-performing students because it could raise anxiety levels for vulnerable learners. A critical issue for low-performing students is to provide instructions that allow them to self-generate ideas about utility value instead of directly telling them the importance (Hulleman et al., 2010b).

High-Performing Students' Disengagement: Control with ITSs

Most teachers can attest to the problematic issues of presented material being too challenging for some while at the same time too rudimentary for others. This dilemma makes learners akin to Goldilocks where the task should not be too hard or too easy. Researchers call this *Goldilocks dilemma* a curvilinear relationship where at the optimum state (inflection point) adding any more or any less of something is undesirable. Researchers discovered evidence indicating a curvilinear relationship between perceived difficulty and boredom; meaning once a task is appropriately challenging adding or subtracting complexity can tip the scale towards negative achievement emotion that results in disengagement (Dettmers et al., 2011; Pekrun et al., 2010). In general, it is more likely high-performing students will have an elevated level of task control (i.e., the task will be easy). Increasing the students' level of control can improve their engagement, but only to a point. Once students believe they have too much control over a task, that it is too easy, it can result in negative achievement emotions (Dettmers et al., 2011) that can result in disengagement (Artino et al., 2012; Schunk, Pintrich, & Meece, 2008). Consider your likely engagement level if you were suddenly transported back to the first grade and needed to spend hours performing addition and subtraction. Student disengagement is a concern with high-performing students in an ITS environment.

Researchers in the ITS community are aware of the challenges these systems can cause with high-performing students (Brenner et al., 2017; Koedinger & Aleven, 2007). As previously mentioned the extra scaffolding provided to learners is particularly helpful to low-performing students (Broza & Kolikant, 2015). The expertise reversal effect suggests the same help that benefits low-performing student can be detrimental to high-performing student success (Salden et al., 2008). A prominent feature of ITSs is to provide *worked examples* to students (Koedinger & Aleven, 2007). A worked example means the solved problem is available to the student and each click for assistance directs their attention to steps of the task before ultimately providing learners with the answer. Early research identified that novice students learned mathematics concepts better and faster when using worked examples compared to unsupported problem solving (Sweller & Cooper, 1985). These worked examples, while helpful to low-performing students, undermine high-performing students (Kalyuga, 2008).

To reduce the expertise reversal effect these worked-out solutions pose, researchers have developed ITS architecture with *fading procedures* (Salden et al., 2008). The idea of a fade procedure is to gradually reduce the number of worked examples (where a completed solution is available) while increasing problems (where learners find the answer; Renkl & Atkinson, 2007; Salden et al., 2008). A significant challenge with this approach is knowing when the transition (fade) should happen. Instructional designers refer to this issue as the assistance dilemma. In addition to this quandary, results from ITSs equipped with fading procedures show only modest learning results (Salden et al., 2008). The positive effects realized have occurred in lab settings, and mainly disappear in classroom settings and, in either case, the positive results vanished quickly (Salden et al., 2008).

A sensible approach to decreasing overly high perceptions of control is simply to make the mathematics problems more difficult. The issue is that learning theorists designed the ITS algorithms to ensure complete mastery of a topic before moving to harder questions or additional content, which is one of the features proposed to contribute to such a large effect size (Foshee et al., 2016; Woolf, 2009; Yang, Gamble, Hung, & Lin, 2014). If high-performers are known to suffer from the expertise reversal effect (Koedinger & Aleven, 2007), and fading procedures have had limited positive results (Salden et al., 2008) is there another approach to dampen adverse effects of control for high-performing students?

A method to consider for high-performing students becoming bored while using an ITS is to instruct them to briefly engage in a cognitively taxing reappraisal. There is flexibility because some scenarios are more difficult to imagine than others. A particularly challenging reappraisal can increase students' cognitive load (Gan, Yang, Chen, Zhang, & Yang, 2017). Research directly related to imagination and cognitive load theory suggests benefits to high-performing students (Cooper, Tindall-Ford., Chandler, & Sweller, 2001). Investigators compared high school students who used worked examples with students who only imagined the worked examples. The former emphasized understanding procedures and concepts while the latter emphasized imagining the processes and ideas. High-performers with experience in algebra benefited from imagination instructions while low-performers with limited experience in algebra appeared unable to use the imagination instruction (Cooper et al., 2001). Researchers suggest the act of imagining the problem requires a more active engagement of the working memory and this facilitates recall in the future (Leahy & Sweller, 2005). Investigators believed requesting students to use their imagination increased cognitive load, and therefore its effect is dependent on the student's level of knowledge (Leahy & Sweller, 2004).

The evidence continues to build that high-performing students suffer performance issues while using an ITS (Brenner et al., 2017; Koedinger & Aleven, 2007). Researchers suggest the problem for these students is the material is not posing enough challenge (Salden, Aleven, Schwonke, & Renkl, 2010; Schwonke, Renkl, Salden, & Aleven, 2011). Additionally, research on reappraisals indicates they are useful in influencing how students appraise a learning task (Gross, 1998, 2014).

In summary, the control-value theory model leads researchers to investigate student appraisals to understand academic outcomes. Students disengagement while using ITS is one such problematic outcome and addressing learner appraisals of control and value regarding the task is appropriate (Pekrun, 2006). ITS developers have done an admirable job in designing tutor systems that maintain a manageable difficulty level (control) for novice students (Woolf, 2009). Since ITSs are handling the challenge level properly, the relevant area to intervene with low-performing students is their value of the topic. There is a well-established body of research to suggest its effectiveness of the utility value intervention on student academic performance (Acee, & Weinstein, 2010; Hulleman et al., 2010b; Menager-Beeley, 2001; Strain & D'Mello, 2015). High-performing students have a distinctly different concern. Researchers have widely documented the issues high-performers have with ITSs reducing their challenge using too much scaffolding (Kalyuga, 2008; Koedinger & Aleven, 2007; Salden et al., 2008). Where the research is lagging, however, is how to correct this issue for high-performing students. The idea of using formative feedback to request students to engage in a cognitively taxing reappraisal with the intention of increasing the perceived difficulty required further testing. Therefore, I conducted a pilot study to understand the intervention and better determine its potential effectiveness for high-performing students. The details of the pilot study and how this influenced the primary study are covered in the next chapter.

CHAPTER 3:

PILOT STUDY

In Chapter 3, I first review the pilot study conducted in the spring of 2017 to investigate whether crucial components of the primary study were feasible. This small study with students at Cedar Valley College in Lancaster, Texas, was done specifically to test a proposed intervention strategy to be eventually used in the primary study. The pilot study is broken into four parts: introduction, methods, results, and discussion. Second, I present the primary study and broadly describe the purpose. The chapter ends by providing the research questions and hypotheses that guided the primary study.

Pilot Introduction

Researchers have identified problems for high-performing student losing interest and becoming disengaged while using intelligent tutor systems (Koedinger & Aleven, 2007). Issues common in ITSs, such as questions being too easy and the pace being too slow, can adversely affect these learners (Salden et al., 2010). Investigators have classified many of these dilemmas under the umbrella term, expertise reversal effect. This effect shows traditional approaches to help low-performing students, such as scaffolding, can have the opposite intended effect and hurt the academic outcomes of high-performers. Methods to fade the scaffolding over time have had limited results (Salden et al., 2008).

Using control-value theory as a model, I proposed lowering high-performing students' sense of control by adding additional challenges to improve engagement and learning for some students. Initially, the intervention to decrease control included two elements to occur within a single condition. The intervention involved observing students and implementing a protocol if the student became disengaged. Once a student became disengaged, I applied two components that occurred in succession.

The first element required students to engage briefly in an off-task behavior (game). The idea of student stopping to play a game was decided because the theory and existing research backed up the proposal. When attempting to improve a student's subjective control, the intervention employed may be as important as how researchers deliver it. From a control-value perspective, allowing students time for an off-task behavior, such as a game, could be valuable if it changes students' perceptions of subjective control. Baker et al. (2011) say an "off-task behavior, within reasonable limits, may actually be beneficial for affect and in turn perhaps even learning" (p. 22). With high-performing students, an off-task behavior in the form of game-based formative feedback may be beneficial for the counter-intuitive reason it lowers their control.

The second element of the intervention used a formative reappraisal (FR) task (students imagining how they would teach the geometry problem they were currently solving). Asking a student to imagine the challenging task of giving instructions was proposed to increase the difficulty for the student and therefore reduce the expertise reversal effect. Leahy and Sweller (2004) have suggested requesting students to use their imagination increases their cognitive load and may only be beneficial for high-performing students. Furthermore, the imagination principle has documented making a task harder improves learning performance for experts (Rey & Buchwald, 2011).

Due to concerns over whether these two factors were both related to a decreased feeling of control, I investigated each part separately using two experimental conditions and compared to a third no treatment condition. There were two concerns with this plan. First, would my proposed intervention increase the challenge and thereby lower a student's perceived control? Second, if it did lower control, would the student see an improvement in any of the key variables of the study: affect, engagement, or learning? I cover the details of conducting the pilot study next.

Pilot Method

In the pilot, I used a quasi-experimental between-subjects approach. I randomly assigned participants to one of three conditions. The three independent conditions were a) FR-Control, b) Game, and c) No Treatment. I measured each student on the dependent variables, and statistical differences between the treatments were investigated using one-way ANOVA, which are described further in the analysis section.

The Office of Regulatory Affairs Human Research Protection Programs at Michigan State University and the IRB committee at Cedar Valley College approved the procedures for this study.

Pilot Participants

The study was completed with 15 students (6 Female, 9 Male) aged 19-26 (M =20.06, SD =2.05). I randomly assigned each student to one of three conditions: FR-Control, game-based, and no treatment condition (five students per cell).

Pilot Apparatus

The students worked individually on the Carnegie Learning ITS, a proprietary system developed by researchers at Carnegie Mellon University. The system provides lessons and then feeds students problems over the content covered. As students work on the problems, the ITS delivers customized feedback. Students completed geometry modules on the following topics: a) “Lines, Rays, Segments, and Angles,” b) “Angles and Angle Pairs,” c) “Angle Properties,” and d) “Introductions to Proofs."

Pilot Measures

This section contains the measures that were used in the pilot study. There were five measures used: Van Hiele Geometry Test, Affect Grid, disengagement, questionnaire, and post-test.

Pilot Van Hiele Geometry Test. The Van Hiele Geometry Test is a 20-item multiple-choice assessment that gauges students' geometry performance (Appendix A). The researcher van Hiele (1980) proposed a stage model for understanding geometry with each stage representing a higher level of geometric thinking. The Van Hiele Geometry Test lists its questions in order of the stages meaning questions get progressively more challenging.

Pilot Affect Grid. The Affect Grid (Russell, Weiss, & Mendelsohn, 1989) is a tool that can allow researchers to quickly measure a subject's affective state (Einöther, Baas, Rowson, & Giesbrecht, 2015). This is a 9 x 9 grid composed of four quadrants, which allows the students to place an X in the appropriate location, permitting the investigator to gauge and score a student's affectual condition (Appendix A). The Affect Grid was printed out, and students ticked the box corresponding to their affective state using a pen.

Pilot disengagement. This measure I obtained by observing the students from another room and counting each time they became disengaged. When the student stopped making progress in the ITS for over two minutes and looked away from the computer for 10 seconds, they were counted as being disengaged. The specific procedure for identifying disengagement I provide in the flowchart in Appendix C.

Pilot questionnaire. The questionnaire used the Patterns of Adaptive Learning Scales (PALS) to measure student's perceived competence (Midgley et al., 2000). I modified the PALS questions to reflect the student task of using the ITS. For example, the questionnaire item, "Even if the work is hard I can learn it" became "Even if the math work in this training is hard, I know I can learn it" (Midgley et al., 2000). The questionnaire had five questions, and the participants responded to each question using a seven-point Likert scale from (1) “not at all true of me” to (7) “very true of me” (Appendix A).

Competence beliefs refer to the appraisals students make regarding their ability to be successful. The control-value theory implies researchers can measure perceived control by students’ perceptions of competence over academic activities and outcomes (Putwain & Symes, 2012).

Pilot post-test. This assessment used 16 questions I developed from the content covered in the modules. Questions were drawn from the material of each module to provide four knowledge acquisition problems per section. These questions were constructed to fit into one of two categories: retention or transfer. The retention category included three types of questions: prompt, assertion, and deep reasoning. Prompt questions were the exact type provided to students by the ITS during the module training. Assertion questions I drew from content explicitly covered by the ITS, but that it did not provide directly to students. Deep reasoning questions were those that require more than recalling facts, but where the students demonstrate causal reasoning and inference to answer. The second category was the transfer of learning questions. I generated questions of this variety from another domain not covered in the modules, but the reasoning to answer them could be found in presented topics. These categories were used to assess the degree to which the material was learned.

Pilot Procedures

Before beginning the study, potential participants read the informed consent (Appendix B). After, each participant took the Van Hiele Geometry Test (Appendix A). This exam was used to determine a student’s broad understanding of geometry. The purpose of the pilot study was to discover the effectiveness of the intervention on high-performing students only. Therefore, I only considered those who scored in the higher range (i.e., high-performers) for the study. A minimum score of 12 was required because it corresponded to the subject being in at least the third stage of Van Hiele’s model.

Next, I matched each qualifying student with two other participants with an equivalent or close to equivalent score meaning each triad had approximately equal broad-based knowledge of geometry. From each triad, I randomly assigned the students with scores of 12 or higher to one of the three conditions: FR-Control, game, and no treatment.

Students then received an overview of how to use the ITS and the expectations for completion. Also, I gave a demonstration on using the Affect Grid (Appendix A). Participants worked alone but could request help if needed. The module took students approximately an hour to complete. While students worked in the ITS, I monitored them in the three conditions and recorded when disengagement occurred (Appendix C). For the FR-Control group participants, I also activated a prompt that requested them to imagine they had to teach the problem they were solving. The prompt included a text box for them to detail their thoughts on the reappraisal. Participants in the game condition were prompted with a short game to complete before being redirected back the ITS. Once students in the three conditions completed the module, they were requested to complete the Affect Grid and the motivation survey.

It took students three rounds to cover an entire geometry unit in the ITS. Upon completion of the third round, students took a post-test.

Pilot Results

The pilot results section is broken into two parts: preliminary analyses and primary analyses. The preliminary analyses helped verify the groups were equivalent. The primary analyses used ANOVAs to examine the effect the participant's condition had on the dependent variables.

Pilot Preliminary Analyses

Before conducting the primary analyses, I first examined whether there were any significant differences between students randomly assigned to the three conditions based on gender, age, and geometry performance. As shown in Table 1 below there were no statistically significant differences between the treatment groups and any variable examined, which suggested these were equivalent groups.

Table 1. *Pilot descriptive statistics of variables in each experimental conditional.*

Variable	**FR-Control**	**Game**	**No treatment**	**Group differences comparison**
Gender	60% M	60% M	60% M	χ^2 (15) = 0.00, p = 1.0
Age	M = 21.60; SD =2.30	M = 21.40; SD = 1.51	M = 23.20; SD = 2.16	F(2,12) = 1.19, p = .34
Van Hiele	M = 12.40; SD =0.55	M = 12.40; SD =0.55	M = 12.8; SD =0.45	F(2,12) = 3.86, p = .40

Pilot perceived competence. An important consideration for running the pilot study was to determine whether the interventions were manipulating experimental conditions as expected. If the intervention worked properly, I expected to see a decrease in students' self-reported perceived competence score for the two conditions: FR-Control and Game.

A one-way ANOVA was conducted with condition (FR-Control, Game, No Treatment) as the independent variable and perceived competence score as the dependent variable. There was a statistically significant main effect across conditions on perceived competence, (FR-Control: M = 5.00, SD = 0.28; Game: M = 5.60, SD = 0.28; No Treatment: M = 5.40, SD = 0.45), F (2, 12) = 3.86, p = <.05.

With a small sample size (N = 15), there was a concern whether the ANOVA test was appropriate because the data set may not have met the basic assumptions needed to run the test with confidence. For additional verification, the nonparametric Kolmogorov-Smirnov (K-S) test was also used to examine the differences between groups. The statistics show only the FR-Control group had a D value = -0.28 where the magnitude exceeded the critical D = 0.27.

Based on the results from the ANOVA and the K-S test, I rejected the null hypothesis and concluded only the FR-Control group had frequencies on the survey distributed differently from the other two groups. Statistically, the FR-Control rated their perceived control lower than the other conditions.

Pilot Primary Analyses

The following analyses examined the effect of condition participants' affect, engagement, and learning. Statistical testing included one-way ANOVAS. Before running the analyses, I set an alpha level of 0.05 as the significance levels for meaningful results.

Pilot affect. A one-way ANOVA was conducted with condition (FR-Control, Game, No Treatment) as the independent variable and Affect Grid score as the dependent variable. There was not a statistically significant main effect of condition on perceived competence, (FR-Control: M = 8.60, SD = 3.05; Game: M = 10.00, SD = 1.58; No Treatment: M = 6.40, SD =2.07), F (2, 12) = 3.86, p = .078.

While the result was not statistically significant, it did show a modest effect size (Cohen's d = 0.83). Also, the group means were higher for both intervention conditions, which confirmed neither of the interventions lowered student affect scores when compared to the no treatment group (an important consideration because there was a concern that lowering control might cause distress and this was not indicated by the affect scores).

Pilot engagement. A one-way ANOVA was conducted with the conditions as the independent variable (FR-Control, M = 5.40, SD = 0.89, Game, M = 5.00, SD = 0.71, and no treatment, M = 6.40, SD =0.89) and student disengagement as the dependent variable. There was not a statistically significant main effect across conditions on the post-test, F (2, 12) = 3.71, p = .056.

Pilot learning. A one-way ANOVA was conducted with the conditions as the independent variable (FR-Control, M = 91.25, SD = 5.60, Game, M = 87.50, SD = 0.00, and no treatment, M = 86.25, SD =5.23) and post-test score as the dependent variable. There not a statistically significant main effect across conditions on the post-test, F (2, 12) = 3.89, p = .22.

While not statically significant, there was a large Cohen's d = 1.08 and over a full standard deviation between the means scores of the control group and the FR group. Additionally, the post-test also had four subsections: prompt, assertion, deep reasoning, and transfer of learning. Because the transfer of learning section was of interest, I ran an additional analysis focusing only on scores for this section.

A one-way ANOVA was conducted with the condition as independent variable (FR-Control, M = 85.0, SD = 12.7, Game, M = 70.0, SD = 20.9, and no treatment, M = 45.0, SD = 27.4) and transfer of learning score as the dependent variable. There was a statistically significant main effect across conditions on the transfer of learning score, F (2, 12) = 3.89, p = .038.

In addition, conducting a K-S test for the transfer of learning section confirmed a statistical difference between the control group and the formative reappraisal group. The largest *D* value found when comparing the FR-Control with the no treatment was 0.80, which exceeded the critical *D* value of 0.60.

Based on the results from the ANOVA and the K-S test I rejected the null hypothesis and concluded only the FR-Control group had frequencies on the survey distributed differently from the other two groups. Statistically, the FR-Control scored higher on this portion of the post-test than the other conditions.

Pilot Discussion

It is uncommon for investigators to make a task intentionally more challenging to elevate performance. Research, however, suggests the approach could work when dealing with high-performing students. Areas such as the redundancy effect, isolated-interactive elements principle, variability principle, and imagination principle all have documented making a task harder improves learning performance for experts (Rey & Buchwald, 2011). The pilot study results indicated a possible intervention that increased the challenge level (i.e., the students reported less perceived competence) to correct for the expertise reversal effect.

The results from the formative reappraisal intervention provided me with three reasons to consider it for addressing student control issues. First, there was a significant decrease in perceived competence for the FR group compared to the no treatment group, indicating a decrease in control. Second, there was an overall increase in the performance of the geometry post-test (over a standard deviation from the control group mean) and a statistically meaningful difference in an important section of the post-test. Third, although students in this group had a more challenging task, their affect scores were not lowered, and the mean score was higher for this group than the no treatment group.

The results of the game condition, however, were problematic. The main issue was the game did not decrease the students' sense of control based on the motivation survey. Since the game was easy (tic-tac-toe), one possible explanation is it was not challenging enough to induce a higher cognitive load. Additionally, while this group had a higher mean average, there was no statistically significant improvement discovered on the test or in any section for the game group. Based on these findings, I dropped the game from consideration in the present study, which I cover in the next chapter.

Pilot Limitations

The small sample size was a limitation of the pilot study. Also, I discovered a flaw in determining a subject's behavioral disengagement. It was possible for a student to be disengaged for a protracted amount of time while other students may have brief disengagements, but more numerous. Take student A who is disengaged for five minutes whereas student B, in that same period, has two brief (30 seconds) disengagements. In this scenario, student A was disengaged for five-minutes and student B for only one-minute yet based on the measurement it would appear student A was twice as disengaged. This statement is unquestionably not the case. I adjusted for this flaw in the primary study by also counting the number of times I considered students disengaged.

The post-test also revealed some issues. No student answered the first 12 questions incorrectly. This situation meant I only discovered differences in the transfer of learning problems. Since these questions represented 25% of the test, it was problematic using the results to make general conclusions. The primary study increased the percentage of transfer of learning problems, and I did beta testing on the retention questions to determine the proper difficulty level before administration.

Pilot Conclusion

Conducting the pilot study gave insights in how to best intervene for high-performing students who become disengaged while using an ITS. Using the knowledge gained allowed improvements to be made for the primary study, which I cover next.

Primary Dissertation Study

The primary dissertation study was built on the pilot study, incorporating the finding into a more extensive exploration of using formative reappraisals (FR) to influence students' perceptions of control and value regarding a learning activity. According to the control-value theory, students who make poor task appraisals will eventually lead to performance issues like disengagement (Pekrun, 2006). The focus of the primary study was to help students improve engagement in the course instruction.

In the primary dissertation study, I investigated three different conditions: FR-Control, FR-Value, and no treatment. FR-Control is a condition where students received a reappraisal intervention intended to influence their perceived control of the academic task. FR-Value is a condition where students received a reappraisal intervention intended to influence their utility value of the academic task. The no treatment group was treated identical to the two FR groups except for the intervention.

In addition to the three conditions, I divided students into two groups: high- or low-performers in the subject (geometry). I then examined differences in these conditions with respect to engagement (affective, behavioral) and learning and considered whether these effects varied as a function of the students' geometry performance group (high and low). Below, I provide more detail about the specific dependent variables included in this primary study.

Engagement Types

Traditionally, engagement refers to the degree with which constructs such as interest, curiosity, motivation, and passion are present in the student, along with "the effort and time they invest and the persistence and resilience they demonstrate towards their goals" (Ge & Ifenthaler, 2017 p. 255). Also, there are different types of engagement (e.g., emotional, cognitive, etc.) This study broke engagement up into two parts, *affective engagement*, and *behavioral engagement* that I define and operationalize below.

Affective engagement. Before defining affective engagement, it is helpful to describe *affect*. Researchers examine two components of the concept affect: *trait-like affect* and *state-like affect* (Linnenbrink, 2006). Trait-like affect is stable and is less likely influenced by a given task, whereas state-like affect tends to fluctuate given the environmental conditions (Linnenbrink, 2006). Using control-value theory as a guide, I opted to investigate state-like affect because of how it is influenced by student appraisals. Conceptually, I define state-like affect as the learner's affective state produced by the interaction between student appraisals and the changing environment as shaped by the task or situation (Linnenbrink, 2006). Operationally, I define affect as modulations in arousal and valence (Pekrun et al., 2010). (Note: valence is the amount of pleasure or displeasure that accompanies a task while arousal is how awake or sleepy a student feels during the work.)

For this study, I conceptually define affective engagement as the student's affective state (happy, bored, confused, frustrated, engaged, etc.), such as in the context of working in an ITS (Ge & Ifenthaler, 2017). Operationally, I define affective engagement as positive valence and moderate arousal along with the satisfaction, or lack thereof, of academic achievement (Pekrun, 2006; Strain & D'Mello, 2015). This operational definition describes a state that reflects similar aspects of Csikszentmihalyi's (1990) concept of *flow* and the decreased sense of passing time that accompanies it. Researchers have associated affective engagement with improved achievement and task persistence (Linnenbrink, 2007; Linnenbrink & Pintrich, 2002; Linnenbrink, Ryan, & Pintrich, 1999).

Behavioral engagement. Conceptually, I use Ge and Ifenthaler's (2017) definition of behavioral engagement as "something we can observe to infer students' persistence, effort, attention, participation, and involvement" (p. 256). The keyword in the definition is *infer*. Technically this variable has no causal connection to a student's internal states. For example, it is possible to show behavioral engagement, yet at the same time be either motivated or not motivated to complete the task. Behavioral engagement is operationalized to mean a student is working on the geometry modules in the ITS.

Learning

Using the literature on schemas, I developed definitions of learning. Conceptually, I define learning as a process where students develop schemas by either modifying or adding new material through assimilation and accommodation and networking the schemas in a way that develops an internal hierarchy of concepts (Piaget, 1976; Rumelhart, 1980). Operationally, I define learning as the students demonstrating they have retained geometry facts and concepts and can draw causal connections between topics.

Research Questions and Hypotheses

RQ1. How do conditions (FR-Value vs. FR-Control vs. no feedback) in an ITS influence student *affective engagement* levels? I hypothesize students should show greater affective engagement in FR-Value and FR-Control conditions than students in the no treatment group (Hypothesis 1a). Moreover, I hypothesize an interaction effect between types of students (low-performance vs. high-performance) and FR conditions: The FR-Value should promote greater affective engagement in low-performance students, whereas the FR-Control should better promote affective engagement for high-performance students (Hypothesis 1b).

RQ2. How do conditions (FR-Value vs. FR-Control vs. no feedback) in an ITS influence student *behavioral engagement* levels? I hypothesize students should show greater behavioral engagement in FR-Value and FR-Control conditions than students in the no treatment group (Hypothesis 2a). Moreover, I hypothesize an interaction effect between types of students (low-performance vs. high-performance) and FR conditions: The FR-Value should promote greater behavioral engagement in low-performance students, whereas the FR-Control should better promote behavioral engagement for high-performance students (Hypothesis 2b).

RQ3. How do conditions (FR-Value vs. FR-Control vs. no feedback) in an ITS influence student *learning*? I hypothesize students should show greater learning in FR-Value and FR-Control conditions than students in the no treatment group (Hypothesis 3a). Moreover, I hypothesize an interaction effect between types of students (low-performance vs. high-performance) and FR conditions: The FR-Value should promote greater learning in low-performance students, whereas the FR-Control should better promote learning for high-performance students (Hypothesis 3b).

CHAPTER 4:

METHODS

This study empirically evaluated the effectiveness of using formative reappraisal (FR) to improve affective engagement, behavioral engagement, and learning in student using an ITS. The research was conducted during a 15-week semester at a community college (Cedar Valley College) located in Lancaster, Texas and consisted of five successive groups each requiring three rounds of data collection. The Office of Regulatory Affairs Human Research Protection Programs at Michigan State University and the IRB committee at Cedar Valley College approved the procedures for this study.

In this study, I used an experimental between-subjects approach. I randomly assigned participants to a 3 (FR-condition) x 2 (high- or low-performance) factorial design (Figure 2). The three independent conditions were a) FR-Control, b) FR-Value, and c) no treatment. Also, I considered students' high- or low-geometry performance. I measured each student on the dependent variables and statistical differences between the treatments were investigated using univariate and multivariate tests, which are described further in the analysis section.

		FR-Condition		
		FR-Control	FR-Value	No Treatment
Geometry Performance	High	Treatment cell 1	Treatment cell 2	No Treatment cell 3
	Low	Treatment cell 4	Treatment cell 5	No Treatment cell 6

Figure 2. The layout of the study design.

The remainder of the chapter details the specifics of how the study was carried out. The methods section is described in seven primary subsections: research design, research questions, participants, apparatus, procedures, measures, and data analysis.

Participants

The participants from the study I drew from the Dallas County Community College District (DCCCD), which is the largest undergraduate system in the state of Texas. It includes seven colleges —Brookhaven, Cedar Valley, Eastfield, El Centro, Mountain View, North Lake and Richland. The student population is diverse: 37.1 percent Hispanic, 24.8 percent black, 8.0 percent Asian, 24.9 percent white, and 5.2 percent all others combined. Most students are attending the institution to earn an associate's degree or the first two years of a bachelor's degree.

Using an a priori power analysis, I determined an appropriate sample size was 113. Cohen (1988) suggests guidelines for a "small," "medium," or "large" effect size and places them at 0.20, 0.50, or 0.80 respectively. A study with a similar intervention and design found an effect size in Cohen's medium range (Strain, & D'Mello, 2015). Setting a medium effect size of 0.5 and $p < .05$ produced an acceptable power of .99 (Cohen, 1988). The study sample ($N = 136$) exceeded the number of participants required to have sufficiently acceptable statistical power.

For this study, there was an emphasis on recruiting students from a variety of mathematics class levels. I met with mathematics faculty members to request they offer the opportunity for their students to join the research experiment. I submitted flyers to mathematics faculty throughout the district with a request to distribute them to students. I offered the study in five two-week periods.

In total, 161 students applied for eligibility in the study. Of these, I excluded 23 students based on their geometry placement or Van Hiele scores. Specifically, the placement results from seven participants indicated they already had prior knowledge of the subject area. An additional 16 students had scores on the Van Hiele Geometry Test that were not a match for the study.

Thus, 138 students were enrolled in the study (63 Female, 75 Male) aged 18-42 (*M* =23.7, *SD* =3.83). Twelve students dropped out of the study after the first round, citing time concerns for why they could not return for subsequent rounds. These students were included in the analyses for round 1; analyses for rounds 2 and 3 were conducted with 126 participants.

Apparatus

The study depended on several devices. The first device was the intelligent geometry tutor that delivered the content students studied during the investigation. The second was the program and equipment that aided in the delivery of the intervention. This section covers the apparatuses used for the study.

The Intelligent Geometry Tutor

The Carnegie Learning ITS was designed to teach multifarious mathematical topics. Unlike some adaptive platforms that inform students when they have made mistakes in their work and continually feeds new problems until students eventually get them right, the Carnegie Learning ITS tells students why they got the problem wrong. The system provides customized feedback with an ongoing formative assessment that provides hints to ensure students receive the immediate support to master mathematics concepts and skills (Figure 3).

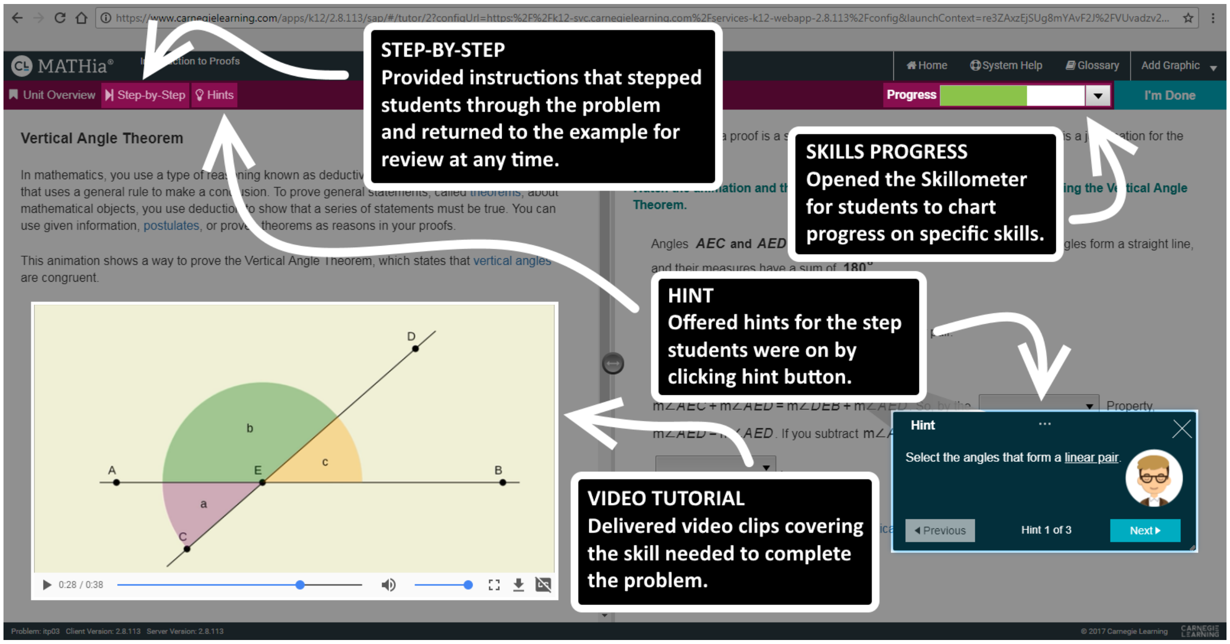


Figure 3. How feedback was delivered to students using the ITS.

The Carnegie Learning ITS was chosen because of an independent study funded by the U.S. Department of Education and conducted by the RAND Corporation (Karam et al., 2017). The two-year research used over 18,000 students in 147 schools throughout seven states. Investigators randomly assigned schools to either a control or experimental group and researchers used *intent-to-treat-analysis*. Results showed mathematics students using the Carnegie Learning ITS nearly doubled gains of a year's worth of learning when compared to the control group (Karam et al., 2017). Evidence from this comprehensive *Gold Standard* study provided assurance the Carnegie Learning ITS was using a valid cognitive model of student learning, which made it an excellent system to investigate an integrative cognitive and emotion model.

The topic of geometry, for use in the study, I selected for two reasons. First, it was unlikely participants were learning about the subject elsewhere as the college was not offering any geometry courses at the time the study occurred. Second, a previous study conducted at the school indicated students felt less apprehensive when working problems using familiar shapes than learning identical concepts in an algebraic format (Xeriland, 2012).

Geometry, in the form of shapes, is often a child's first introduction to mathematics. The field is sufficient that it can carry a student through progression in their educational progress—from kindergarten to advanced graduate studies. One of the most important concepts covered in geometry is the two-column proof. This topic teaches logical reasoning and deduction that are central to understanding mathematics. Lucast (2003) proposed the proof is fundamental for problem-solving because it requires the same cognitive processes.

When students first logged into the system, they completed a tutorial on how to solve problems for the current unit (Figure 4). The ITS briefed students on the outline of the content covered in the module. At the completion of the tutorial, students took a learning check to ensure they were ready to proceed.

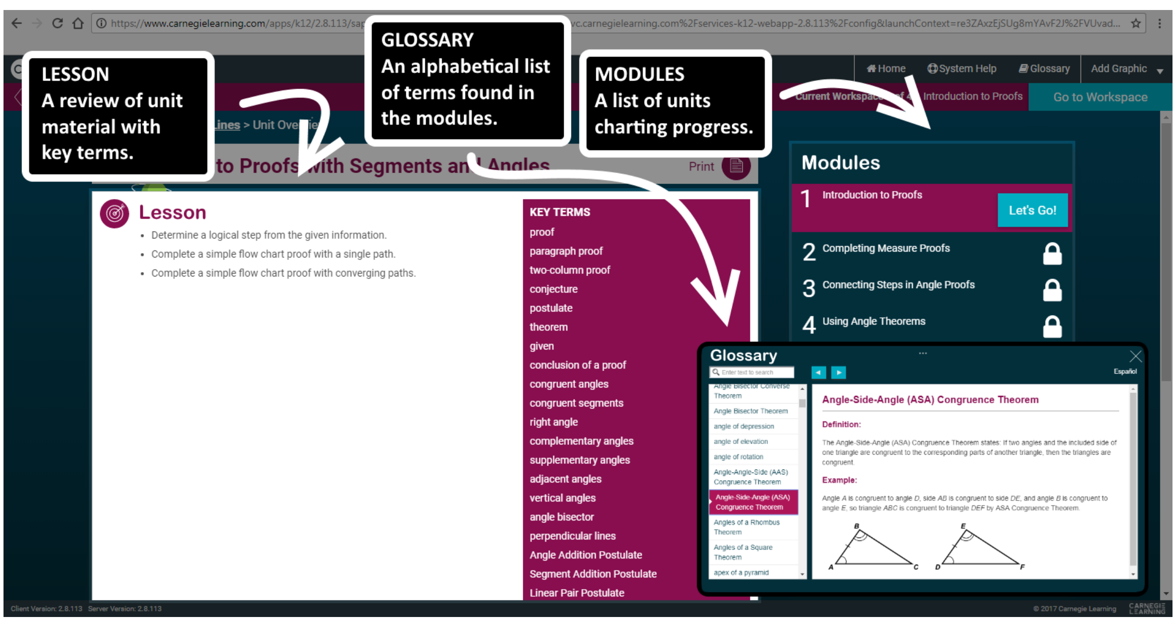


Figure 4. Module tutorial for students.

After the completion of the review, students worked individually on geometry modules that included the following four titles:

1. “Lines, Rays, Segments, and Angles” covered identifying and naming lines, rays, segments, and angles and instructed students on how to work with measures of segments and angles.
2. “Angles and Angle Pairs” covered measuring and classifying angles.
3. “Angle Properties” instructed students on how to calculate and justify angle measures.
4. “Introductions to Proofs" instructed students how to derive logical steps from given information and how to complete simple flow chart proofs involving single and complex paths.

An example of the Introduction to Proofs module is in Figure 5. The ITS incrementally guided students through tutorials with an end goal of being able to write proofs of statements in Euclidean geometry.

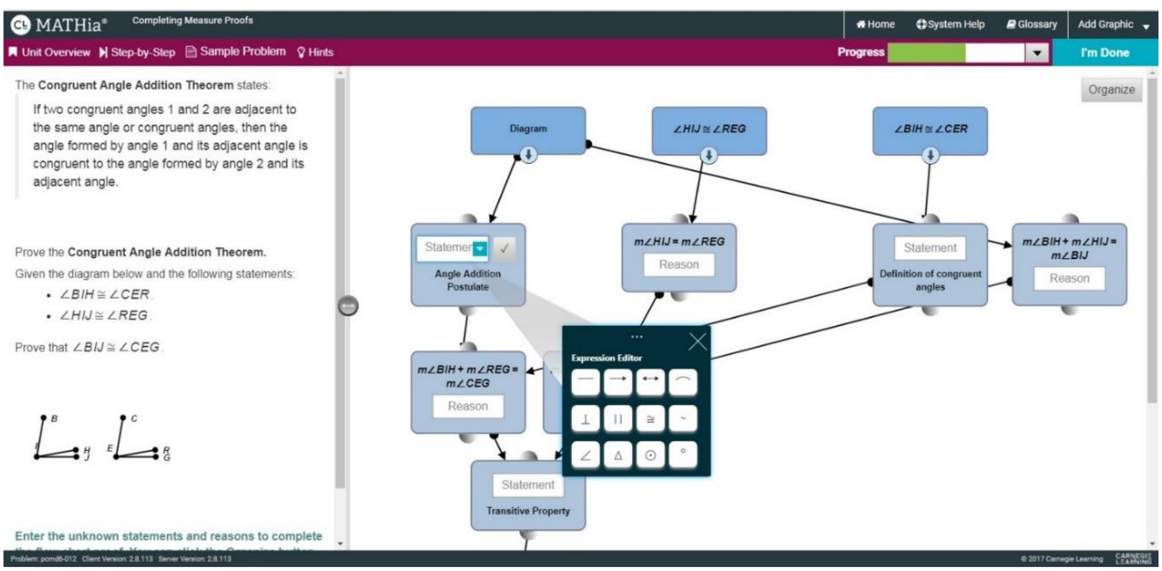


Figure 5. Carnegie Learning Geometry Tutor.

Intervention Delivery Tools

A delivery tool incorporated in this study was a Dell Inspiron 17 5000 computer with built-in Webcam. The laptop was setup with Skype version 7.38.0.101 and a screen sharing program called join.me ©. On the computer, students used the proprietary commercial system known as the Carnegie Learning ITS®. As described in greater detail in the section below, each observer executed an additional program during ITS use that prompted the participants with the interventions and captured their responses (Figure 6). The add-in program was created using Visual C++ 2017 specifically for the study, and I stored student responses in Access 2016.

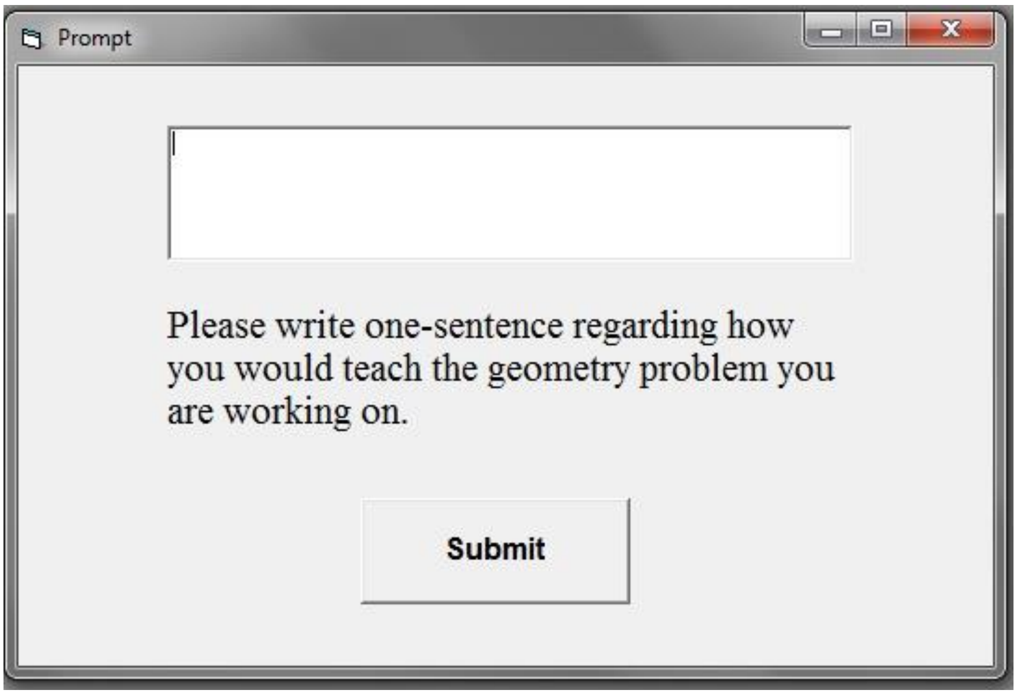


Figure 6. Program for prompting the participants.

Procedures

Students were assessed on their initial motivation and prior knowledge and then completed an ITS geometry module (Round 1). After, students scheduled a time to come back and finish an additional geometry module (Round 2). Lastly, students completed the final ITS geometry module and took a post-test (Round 3).

The study began with informing potential participants about the general aim, duration of the investigation, and reading the informed consent (Appendix B). Next, students took the geometry placement test (Appendix A). The evaluation assessed the four topics in the ITS modules. The geometry placement assessment was only used to eliminate any student who already has substantial knowledge of the material. This assessment helped verify the students had limited prior knowledge of the task. The reason for this confirmation is it increased the experiment's external validity by ensuring participants resembled learners who would normally study the topic.

Once the students qualified for the study, each participant took the Van Hiele Geometry Test. The exam assessed a student's broad understanding of geometry but was not limited to one area. The Van Hiele Geometry Test differed from the geometry placement test because the first was wide-ranging and the second was narrowly focused. I chose the Van Hiele test because it assessed students broad understanding of geometry or referred to hereafter as geometry performance. The research questions separate students with high and low geometry performance and this test is how I determine that distinction.

Using the first 30 students (group 1), I obtained statistics on the Van Hiele test scores (M=9.90, SD=1.97). I provide the Van Hiele frequency distribution for group 1 in Table 2. I used this information to develop a split between those classified as having high geometry performance and those classified as having low geometry performance. I placed students with Van Hiele Geometry Test score above ten in the high-performance group. Those with a score below ten, I put in the low-performance group. Students who achieved a score of ten did not definitively fall into one of the two groups and therefore did not move forward in the study.

Table 2. *Frequency distribution of Van Hiele scores for group 1.*

Score	Frequency (*f*)
13	4
12	5
11	5
10	3
9	5
8	5
7	3

By matching each student with two other participants with an exact score meant each triad had equivalent broad-based knowledge (performance) of geometry. No student started the study until there were two others with the exact same score. This approach prevented students with outlier scores from potential skewing the results. From each triad, I randomly assigned the students with scores higher or lower than 10 to one of the three conditions.

After random assignment, I gave them an overview of how to use the ITS and the expectations for completion. Also, students received instructions on how to complete the Affect Grid (Appendix A) at the end of each unit. Participants worked in a setting alone, but with the ability to press a button to get help from the experimenter in the case of a technical issue. After the students finished the prescribed unit (lasting approximately an hour to an hour and a half), they took the Post-Session Engagement questionnaire and completed the Affect Grid. Finally, I gave subjects the motivation survey.

Following the first session, the students made an appointment to come back and finish more units in the ITS, following similar procedures as those used for the first round. To complete the study, students needed to finish three rounds, which allowed enough time to get through the entire section of the ITS. Once students completed their three sessions, they took a post-test. This overview covered the setup for each condition in the study, including the control group. The next sections include how I manipulated the independent variables.

Wizard of Oz Experiment

The Wizard of Oz (WOZ) experiment has become common in experimental psychology and usability engineering (Rosner et al., 2015). The WOZ approach is an experiment where the subjects believe they are interacting with a computer system that is autonomous, but at some point, a member of the experimentation team takes the role of the machine. For example, a subject may assume a computer can read levels of frustration in a student and act accordingly when the intelligence behind this is the experimenter in another room observing and serving on the machine's behalf.

The WOZ experiment was an appropriate fit for this study to aid me in screening the effectiveness of testing the interventions. By using an observer (wizard) to detect disengagement, it eliminated technological challenges of rigging an automatic detection unit. Although, there were existing computerized detection mechanisms built for ITSs; therefore, adding the missing element provided by the observer was entirely feasible in future applications of the intervention.

Using the WOZ experiment allowed an outside observer to track students' gaze patterns and their computer screens. Once disengagement occurred (see Appendix C for details), the observer could implement the intervention. Considering the number of participants in the study (N = 138) and because students could drop-in to work on a module throughout the day, two individuals served as the observers to detect student disengagement. It was necessary to familiarize themselves with when to consider the student disengaged. The observers jointly monitored the first five students to verify they both agreed on what was considered a disengagement before watching the learners independently.

No Treatment Condition

Everything that occurred in this condition applies to participants across the study. The only difference was the no treatment group did not receive an FR strategy.

All students in the study began with a guided learning session provided by the ITS. The guided learning tutorial gave learners a step-by-step procedure on how to solve problems for the unit. The lesson included a brief outline of the module to prepare students for what was to follow. The system added a learning check with practice problems allowing students to verify their readiness.

Following the overview of the unit, students in this condition began work on the geometry module. Once the student started the ITS module one of the observers outside of the room served as the disengagement detector by remotely monitored both the student and their computer screen. Participants were not made aware external monitoring was occurring. Since only one student went through at a time it reduced the possibility of cross-contaminations where participants might see their condition was in some way different from other students.

The FR-Control Condition

In the FR-Control condition, after completing their unit overview, the observer activated the following student writing prompt.

> Type a short essay (1-3 paragraphs in length). Picture yourself as a college professor that teaches this topic to first-year students. Describe what techniques you would use to improve student understanding of the subject. You'll probably need more practice with the method to appreciate how to instruct students, but for purposes of this writing exercise, please focus on how this technique could be taught to college students and give examples.

After completing their essay, students began work with the ITS geometry modules.

If either observer considered the student disengaged from the ITS (see Appendix C), the resulting action was an activated dialog box with instructions to imagine they were teaching the current problem to a student. A prompt instructed the participant to write a one-sentence response on how they would explain the problem they were working on to college students (Figure 6). The writing reappraisal process was a brief interruption that sought to reinforce the first intervention had it worn off.

After the FR prompt, the student was redirected back to the ITS module to continue with the geometry module. Using formative feedback best practices, the observer gave a maximum of five interventions during any a session and a minimum of 10 minutes transpired between responses to avoid them from becoming too burdensome and defeating the purpose (Shute, 2008).

Before beginning sessions two and three, students in this condition reviewed the essay they wrote at the start of the study. After reading the section, students received the following prompt: “Now that you have covered more of the material on this topic, and have a stronger grasp of the subject, is there any additional information you would add to your essay.” Once students completed this prompt, they began their new module and again receive FR if they became disengaged.

The FR-Value Condition

The intervention in this condition combined utility value used by Hulleman et al. (2010b) and cognitive reappraisals employed by Gross and Thompson (2007). A standard approach to reappraisal interventions is to ask the subject to imagine the task from another person's perspective (Goldin et al., 2008). Participants in this group imagined they were advisors to first-year students. Before proceeding to the start of the geometry module, the observer activated a dialogue box with a writing assignment. Using a utility value writing prompt developed by Hulleman and colleagues (2010b), I modified theirs to better fit the current study:

> Type a short essay (1-3 paragraphs in length). Picture yourself as a college advisor and briefly describe the relevance of this topic to a first-year student. You'll probably need more practice with the technique to appreciate its significance, but for purposes of this writing exercise, please focus on how this material could be useful to college students and give examples.

After completing their essay, students began work with the ITS geometry modules.

If the observer considered the student disengaged from the ITS (see Appendix C), the resulting action was an activated dialog box with instructions to imagine they were a college advisor working with college students. The system prompted them to write one sentence on what an advisor would likely tell a first-year student is the relevance of the work they are doing. This writing reappraisal process was a brief interruption that sought to reinforce the first intervention had it worn off.

Immediately after the students typed their sentence into the text field, the subjects were redirected to the ITS to continue with the geometry module. The observer intervened a maximum of five times during a session and at least ten minutes needed to transpire between interventions. I chose this procedure in accord with formative feedback best practices (Shute, 2008).

At the start of sessions two and three, students reviewed the essay they wrote at the beginning of the study. Once students read their previous submission, they received the following prompt: "Now that you have covered more of the material on this topic, and have a stronger grasp of the subject, is there any additional information you would add to your essay."

Upon completion of this prompt students began their new module and again receive FR if they became disengaged.

Measures

To measure key variables, I used several assessments: pre-test, a survey, a questionnaire, manipulation checks, and a post-test. Figure 7 displays the different tools used for taking measures and at what point during the study they occurred. I describe the instrumentation below.

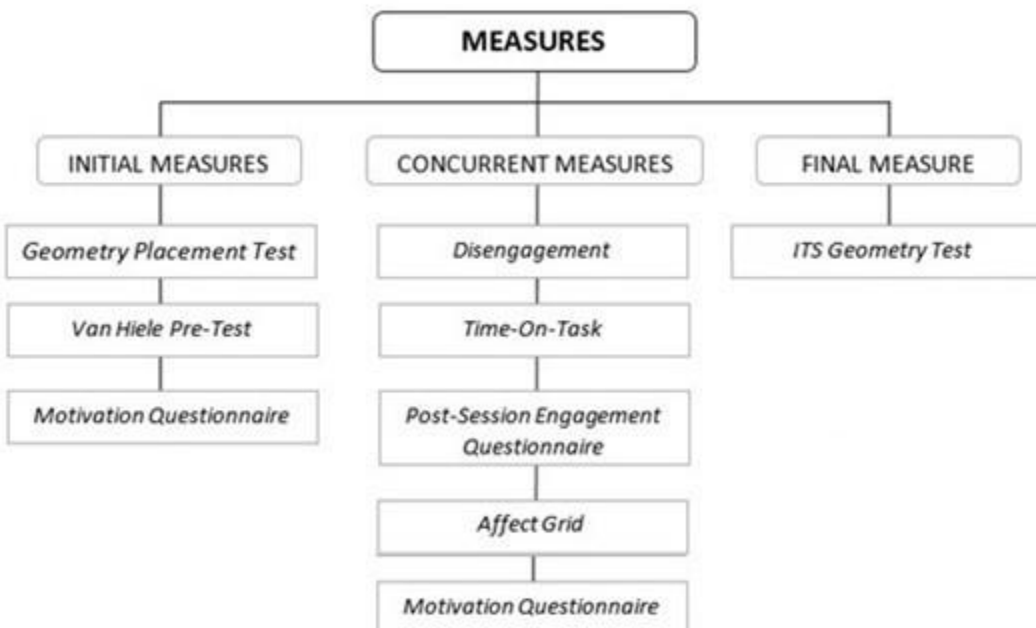


Figure 7. Overview of measures and when they occurred during the study.

Initial Measures

This section covers the instrumentation I used in the study before students started the geometry modules in the ITS.

Geometry placement test. This five-question test, administered by hand, covered material students would later study in the four geometry modules (Appendix A). I constructed it as a screening tool that used fundamental questions a student already educated in the topic areas would know. Any student who answered more than two of the five questions correctly did not move forward in the study; therefore, I removed seven students from consideration based on their test score here. I chose the cutoff score because it was extraordinarily unlikely to score above 40% by guessing. Any score above 40% assured me the participant had at least some knowledge in the area and was not an appropriate candidate for the study. This instrument served as a filter for students who would not usually take a course covering this material.

Van Hiele pre-test. The Van Hiele Geometry test is a 20-item multiple-choice assessment that was developed to test "the ability of the van Hiele theory to describe and predict the performance of students" (Usiskin, 1982, p. 8). The version used is the only van Hiele instrument available for group administration, which has standardized response choices (Crowley, 1990). The Van Hiele Geometry test has reliably served as a tool in many research settings (Fuys, Geddes, & Tischler, 1988; Senk, 1989; Usiskin, 1982).

According to van Hiele (1980), there are four levels of understanding in geometry: recognition, analysis, order, and deduction. These levels describe the way students reason about shapes and other geometric ideas. The van Hiele levels are not age dependent and are more related to student past experiences (Burger & Shaughnessy, 1986). Additionally, the levels are sequential, and students are required to pass through one before arriving at the next (Burger & Shaughnessy, 1986).

The Van Hiele Geometry test is composed of four subtests, with five questions each. The subtest items correspond with cognitive abilities believed to exist at each level (Usiskin, 1982). Using a sample of 2,699 students from five states enrolled in one-year geometry courses, Usiskin (1982) reported the Van Hiele test had a .69 correlation to the Comprehensive Assessment Program Geometry test, which is widely used in schools to assess student knowledge of geometry. Additionally, researchers discovered students at each Van Hiele level were more successful in learning proofs when compared to students in proceeding levels (Senk, 1989). Researchers found that the test could predict the probability of proof writing success with up to 87% accuracy (Usiskin, & Senk, 1990).

The Van Hiele Geometry test was not a pre-test that matched to a post-test. Instead, it measured a student's model of geometric thinking for the purposes of determine high and low performers. A typical geometry test requires memorizing facts before taking it, whereas for the Van Hiele Geometry Test it is doubtful memorization would benefit since the test gauges a student's basic geometry understanding to be successful in the next level. For example, researchers find students struggle to learn concepts in the final stage, deduction, when their test results do not indicate mastery in the first three levels (Usiskin, 1982). The full test is available from University of Chicago School Mathematics Project, and I provide sample questions for determining each Van Hiele level in Appendix A.

Pre-motivation questionnaire. This 11-question motivation questionnaire I developed from two well-established surveys: Patterns of Adaptive Learning Scales (PALS; Midgley et al., 2000) and Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, Smith, Garcia, & McKeachie, 1991). Specifically, the survey measured perceived competence (Cronbach's $\alpha = .81$ for five items) and task value (Cronbach's $\alpha = .88$ for six items). Conceptually perceived competence is the student's belief in the likelihood of their success in a task.

The questions from PALS and MSLQ were slightly modified to fit the ITS environment. For example, the survey item, "Even if the work is hard I can learn it" became "Even if the math work in this training is hard, I know I can learn it" (Midgley et al., 2000). Participants responded to each question using a seven-point Likert scale from (1) "not at all true of me" to (7) "very true of me" (Appendix A).

Concurrent Measures

Instrumentation in this section I used concurrently with the students working on the geometry modules with the ITS. My assistant or I took the following measures during the time students were working with the geometry modules.

Disengagement. For this measure, the observer recorded the number of times a student became disengaged from the ITS. What constituted a disengagement was defined by using the flowchart in Appendix C. Note: the FR intervention only occurred after ten minutes elapsed from the previous occurrence, but the observer tallied disengagements regardless of the amount of time between them.

Time-off-task. An important variable was time-off-task. The pilot study did not use this variable, and it rendered the disengagement results inconclusive. To correct the issue and reduce the influence of a confounding variable for disengagement, the observer measured time-off-task by starting a stopwatch each time the student became disengaged from the ITS. The time only stopped once the student moved the mouse to continue work on the geometry problem. At the end of each round, the observer recorded the total time the student was off task.

Post-session engagement questionnaire. This three-question survey was administered electronically at the end of each geometry module to assess students' engagement levels. Participants reported their engagement level at the beginning, middle, and end of the session (e.g., how engaged were you when you started today's session). Students ranked this on a seven-point Likert scale from (1) very bored to (7) very engaged. A similar study used this questionnaire (D'Mello et al., 2012). I administered the post-session engagement questionnaire at the end of each tutoring session (round 1 $\alpha = .91$ for three items, round 2 $\alpha = .89$ for three items, and round 3 $\alpha = .91$ for three items).

Affect grid. The Affect Grid is a single item scale designed to measure affect along dimensions of pleasure–displeasure and arousal–sleepiness (Russell et al., 1989). This chart asked the student to place an X on a 9 x 9 grid that directly measures valence and arousal (see Appendix A). The *x*-axis measures valence (ranging from very negative on the far left to very positive on the far right) and the *y*-axis measures arousal (ranging from very low arousal at the bottom to very high arousal at the top of the grid; Larcom, & Isaacowitz, 2009). I instructed the participants to tick the box that best corresponded to their affect state after completing the geometry module. The benefit of the Affect Grid was it allows subjects to quickly report an affective state without having to use an affective label (Strain & D'Mello, 2015). I administered the Affect Grid at the end of each tutoring session (round 1 $\alpha = .88$ for two items, round 2 $\alpha = .84$ for two items, and round 3 $\alpha = .77$ for two items).

Motivation questionnaire. Students completed the 11-item motivation questionnaire (previously mentioned above as pre-motivation questionnaire) at the end of the three sessions. The questionnaire addressed students' perceived competence (round 1 $\alpha = .78$ for five items, round 2 $\alpha = .84$ for five items, and round 3 $\alpha = .80$ for five items) and task value (round 1 $\alpha = .89$ for six items, round 2 $\alpha = .91$ for six items, and round 3 $\alpha = .92$ for six items).

At the beginning of the study, before students started with the ITS, the survey questions were phrased in the present tense. For instance, a question on the pre-motivation questionnaire read, "I think the material in this training will be useful for me to learn." The motivation questionnaire done after starting the sessions had questions in the past tense. For example, the same question above was phrased, "I think the material in this training was useful for me to learn." Applying the survey twice was an additional control for potential differences in motivation prior to the random assignment.

The questionnaire served as a manipulation check for the two experimental interventions FR-Value and FR-Control. Having students take the survey at the end of each session allowed me to verify the intervention was manipulating the proper condition at each stage of the experiment. Also, this method allowed me to validate signs of the manipulation effectiveness growing or shrinking throughout the study.

Final Measure

Once the student completed the third session on the ITS, I used the following measures.

ITS geometry post-test. This 8-question test covered content found in the modules (Cronbach's α = .87). I designed the ITS Geometry Test to measure student learning of the geometry concepts taught by the ITS. The test assessed student knowledge of writing two-column proofs included in the final module. I considered these proofs a quality indicator of overall knowledge gained since the purpose of each module was to build up the students' understanding of constructing them. I categorized half of the test as retention questions (Cronbach's α = .78 for four items). These items should have looked familiar to the students because they were variants of queries raised during the ITS trials. The retention questions were further broken down into recognition and recall subcategories.

Educators often assess two types of student memory retrieval: recognition (e.g., multiple-choice) and recall (e.g., fill-in-the-blank). The results of the pilot study indicated writing two-column proofs was challenging for students and therefore testing both recognition and recall allowed the researcher to look for differences in memory activation levels. On the post-test recognition is assessed with multiple choice questions, whereas recall required students to fill in the blank.

The other half of the exam used the transfer of learning questions (Cronbach's $\alpha = .80$ for four items). Researchers describe transfer of learning as the ability to apply what one has learned in one situation to a different situation. (Rebello, Cui, Bennett, Zollman, & Ozimek, 2007). The ITS Geometry Test distinguished between two types of transfer of learning: near and far. The *near transfer* is when a student can take a learned concept from a domain and apply it to a new context of that same domain (Rebello et al., 2007). On the post-test students were asked to solve geometry proofs using the same theorems and postulates they studied but apply it to geometry figures students did not cover during the ITS training. For example, if they learned to use the substitution property to solve a proof using line segments the transfer of learning question (classified as near) in turn required students to use the substitution property to solve a proof using triangles. The *far transfer* is when a student can take a learned concept from one domain and apply it to a new domain (Rebello et al., 2007). For instance, on the post-test students were required to solve proofs that needed the same logic they used in geometry, but now needed to be applied to algebraic concepts to solve the problem.

The transfer questions required more than a recall or recognition, and the students needed to demonstrate causal reasoning and inference to answer them correctly (see Appendix A for the full test). I administered the ITS Geometry Test at the end of the study.

Data Analysis

The study used a randomized block design where I divided subjects into relatively homogeneous subgroups or blocks based on their pre-test scores before randomly assigning them to one of three conditions: 1) students who received only FR-Control, 2) students who received only FR-Value and 3) students who received no treatment. An ANOVA was used to verify the mean van Hiele scores did not significantly differ across the three groups. To test whether the FR treatment was producing the intended effect for the relevant condition, I conducted an ANOVA to examine differences across conditions for perceived competence or task value assessed at each round. Specifically, FR-Control was expected to influence perceived competence whereas FR-Value was expected to influence task value when compared to the other conditions.

For the main analyses, I examined the effects of FR intervention using a series of 3 (FR conditions) x 2 (initial level of geometry performance) ANOVAs and MANOVAs for affective engagement (rounds 1 through 3) and learning (after round 3). When conducting an ANOVA, there are four assumptions to consider: independence, scale of measurement, normality, and equality of variance. Independence was assumed because the study used a randomized design. The instruments in the study were scale and not ordinal. The skewness and kurtosis for the data analyzed for each group were between 1 and -1 suggesting the data was normally distributed. Finally, I ran Levene's tests for the data distributions, and none were significant therefore homoscedasticity was assumed.

Affective Engagement Analysis

There were two instruments for this variable: Affect grid and Post Engagement Questionnaire. Together these measures were combined to measure effective engagement. Past researchers (D'Mello et al., 2012) used this same approach to quantify engagement. First, I standardized the variables obtained by creating *z*-scores. The sum of these scores was the measure of affective engagement.

A 3 x 2 ANOVA to examine the effect of the condition on affective engagement between the experimental groups. An ANOVA for main effects and interactions examined change in effectiveness between levels of expertise and the three treatment conditions.

Behavioral Engagement Analysis

The measurements used for this variable were time-off-task and number of disengagements. To conduct the analysis, I used a MANOVA on the main effect of the condition on the measurements: time-off-task and number of disengagements. Additionally, two univariate ANOVAs explored the effects of the condition on the measurements separately.

Learning Analysis

The instrument used for this variable was the Carnegie ITS Geometry Test. First, a 3 x 2 ANOVA examined the effect of the conditions on the test scores between the groups and looked for an interaction between levels of expertise across the three experimental conditions. I also used additional ANOVAs to analyze the two subsections of the test: retention and transfer of learning.

CHAPTER 5:

RESULTS

Results are presented in two sections. The first reports the preliminary analyses, which examined the equivalence of the participants in each condition and performed a manipulation check assess the effectiveness of the intervention. The second presents the primary findings, where the three dependent variables are investigated with a series of analyses of variance, either ANOVA and MANOVA.

Preliminary Analyses

Prior to conducting the primary analyses, I first examined whether there were any significant differences between students randomly assigned to the three FR conditions using gender, age, and geometry initial performance, as well as initial motivation (task value, perceived competence). As shown in Table 3, Chi-square and analysis of variance indicated no statistically significant differences between the treatment groups or any variable examined.

To assess the effectiveness of the intervention on the students' affective engagement, behavioral engagement, and learning I ran a series of one-way ANOVAs to confirm the experimental conditions functioned as expected for each session. I sought to manipulate the dependent variable in two of the conditions: FR-Control and FR-Value. For the FR-Control manipulation, I expected to see students' self-reported perceived competence lowered while for the FR-Value condition the manipulation should have elevated students' reported task value in the subject. Significant results were examined using the post hoc test least-square-differences (LSD).

Table 3. *Descriptive statistics of variables in each experimental conditional.*

Variable	**Condition**			**Group differences comparison**
	FR-Control	**FR-Value**	**No treatment**	
Gender	52% M	60% M	52% M	χ^2 (138) = 0.58, *p* = .75
Age	*M* = 23.67; *SD* =4.35	*M* = 24.53; *SD* = 3.88	*M* = 23.68; *SD* = 3.21	*F*(2,135) = 0.75, *p* = .47
Van Hiele	*M* = 10.09; *SD* =2.21	*M* = 9.82; *SD* =2.27	*M* = 9.94; *SD* =2.19	*F*(2,135) = 0.16, *p* = .85
Task Value	*M* = 4.24; *SD* =0.73	*M* = 4.13; *SD* =0.69	*M* = 3.95; *SD* =0.85	*F*(2,135) = 1.61, *p* = .20
Perceived Competence	*M* = 3.95; *SD* =0.60	*M* = 4.16; *SD* =0.58	*M* = 4.01; *SD* =0.54	*F*(2,135) = 0.50, *p* = .61

Perceived Competence

The purpose of the perceived competence measure was to verify FR-Control intervention lowered the perceived competence score of subjects in the FR-Control condition compared to the other conditions. The lowering of the perceived competence was the aim of the manipulation to address control issues and expertise reversal effect. Specifically, I expected the score to produce a main effect for the condition. Figure 8 below uses a bar graph to show the mean differences between the three conditions across the three rounds.

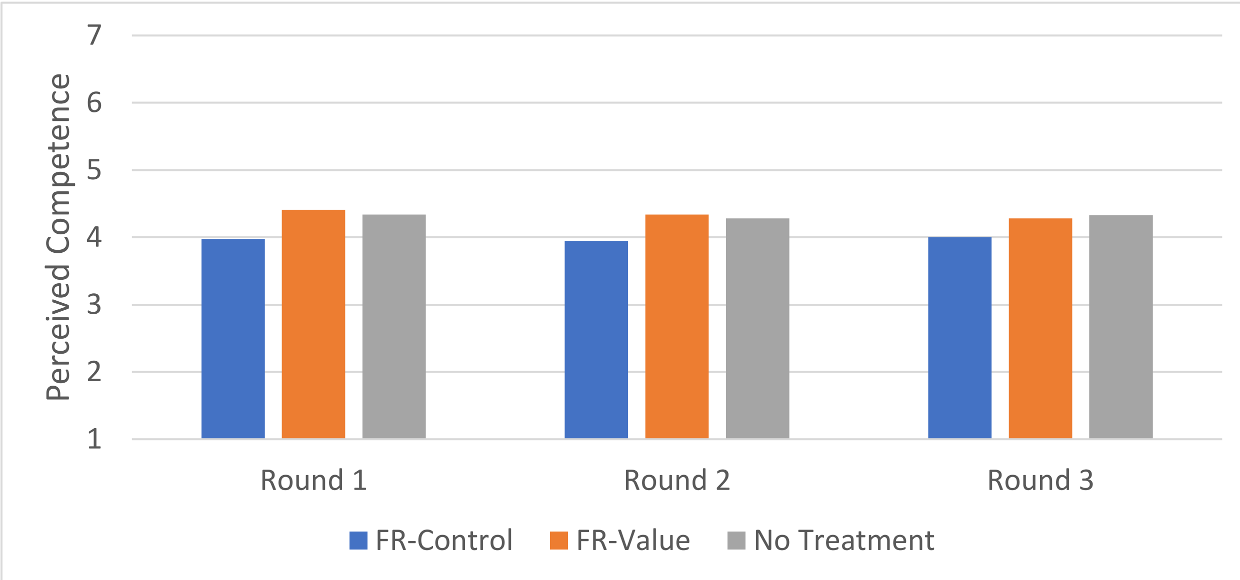


Figure 8. Perceived competence mean average scores across three rounds; scale ranged from 1 "not at all true" to 7 "very true."

Three one-way ANOVAs were conducted for the FR condition as the independent variable (FR-Control, FR-Value, and no treatment) and perceived competence as the dependent variable for rounds 1, 2, and 3, respectively. There was a statistically significant main effect across conditions on perceived competence for round 1, $F(2, 135) = 4.32$, $p = .015$, and round 2, $F(2, 123) = 3.42$, $p = .036$, and round 3, $F(2, 123) = 3.08$, $p = .049$. Additionally, the means and standard deviations across conditions I summarize in Table 4.

Table 4. *Mean averages and standard deviations of perceived competence scores by condition and initial performance for three rounds.*

FR	Round 1		Round 2		Round 3	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
FR-Control	3.98	0.80	3.95	0.76	4.00	0.73
FR-Value	4.41	0.74	4.34	0.69	4.35	0.70
No Treatment	4.34	0.68	4.28	0.73	4.33	0.72
Total	4.24	0.76	4.19	0.74	4.23	0.73

*Note: The measure of perceived competence ranged from 1 "not at all true" to 7 "very true".

Post hoc testing using LSD comparisons indicated, during round 1, participants in FR-Control had significantly lower perceived competence scores than the no treatment group (p = .023) and the FR-Value group (p = .007). During round 2, participants in FR-Control had significantly lower perceived competence scores than the no treatment group (p = .043) and the FR-Value group (p = .016). During round 3, participants in FR-Control had significantly lower perceived competence scores than the no treatment group (p = .038) and the FR-Value group (p = .030). These finding offered compelling evidence that the FR-Control invention successful manipulated perceived control as intended.

Task Value

The measure of task value was used to confirm the FR-Value intervention raised the scores of participants in that condition compared to the other conditions. The manipulation would be confirmed if the score to produce a main effect for the conditions. Figure 9 graphs the mean differences between the three conditions across the three rounds.

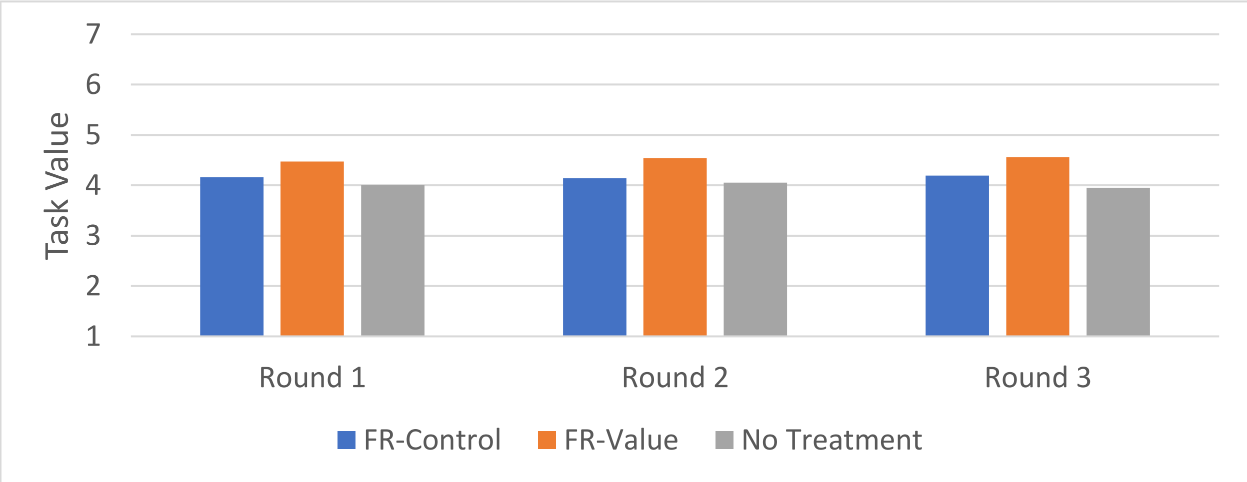


Figure 9. Task value mean average scores across three rounds; scale ranged from 1 "not at all true" to 7 “very true.”

The means and standard deviations across conditions are summarized in Table 5 below. Three one-way ANOVAs were conducted for the FR condition as the independent variable (FR-

Control, FR-Value, and no treatment) and task value as the dependent variable for rounds 1, 2, and 3, respectively. There was a statistically significant main effect across conditions on task value for round 1, $F(2, 135) = 3.06$, $p = .050$, and round 2, $F(2, 123) = 3.76$, $p = .026$, and round 3, $F(2, 123) = 3.93$, $p = .008$.

Post hoc testing using LSD comparisons indicated, during round 1, participants in FR-Value had significantly higher task value scores than the no treatment group ($p = .017$) but did not significantly differ from the FR-Control group ($p = .10$). During round 2, participants in FR-Value had significantly higher task value scores than the no treatment group ($p = .012$) and the FR-Control group ($p = .041$). During round 3, participants in FR-Value had significantly higher task value scores than the no treatment group ($p = .003$) but did not significantly differ from the FR-Control group ($p = .058$). These finding offered compelling evidence that the FR-Value invention successful manipulated the condition**.**

Table 5. *Mean averages and standard deviations of task value scores by condition and initial performance for three rounds.*

FR	Round 1		Round 2		Round 3	
Condition	*M*	*SD*	*M*	*SD*	*M*	*SD*
FR-Control	4.16	0.98	4.14	0.91	4.19	0.88
FR-Value	4.47	0.85	4.54	0.87	4.56	0.86
No Treatment	4.01	0.91	4.05	0.84	3.95	0.90
Total	4.21	0.93	4.24	0.89	4.23	0.91

*Note: The measure of task value ranged from 1 "not at all true" to 7 “very true”.

Primary Analyses

The following analyses examined the effect of condition and initial geometry performance on subjects’ affective engagement, behavioral engagement, and learning. Statistical testing included between-subjects 3 (FR-Control, FR-Value, no treatment) x 2 (high/low Initial Geometry Performance) using ANOVAs and MANOVAs. Before running analyses, an alpha level of 0.05 was selected for significance levels of results. I was specifically interested in the comparisons between the FR conditions and the no treatment and whether there was an interaction between condition and initial geometry performance. I hypothesized that the FR treatments would improve the three dependent variables in the following ways, 1) affective engagement would have higher scores for the Affect Grid and post-session engagement questionnaire, 2) behavioral engagement would have lower scores for disengagement and time-off-task, and 3) learning would have improved scores on the post-test. Finally, I hypothesized results by conditions would vary depending on student initial geometry performance—low performers would benefit most from FR-Value while high performers would benefit from FR-Control.

Affective Engagement

The first research question asked whether the FR conditions differentially predicted students' affective engagement. An ANOVA using the FR condition (FR-Control, FR-Value, and no treatment) and initial geometry performance (high and low) as independent variables and affective engagement as the dependent variable was conducted for rounds 1, 2, and 3, respectively. For these analyses, I expected there would be a main effect of the FR condition compared to the no treatment group. Also, I expected students in the FR-Value who were low-performance would show greater improvement than high-performance students. I expected the opposite for FR-Control where high-performance students were thought to see the most benefit. Figure 10 graphs the mean differences between the three conditions for round 1.

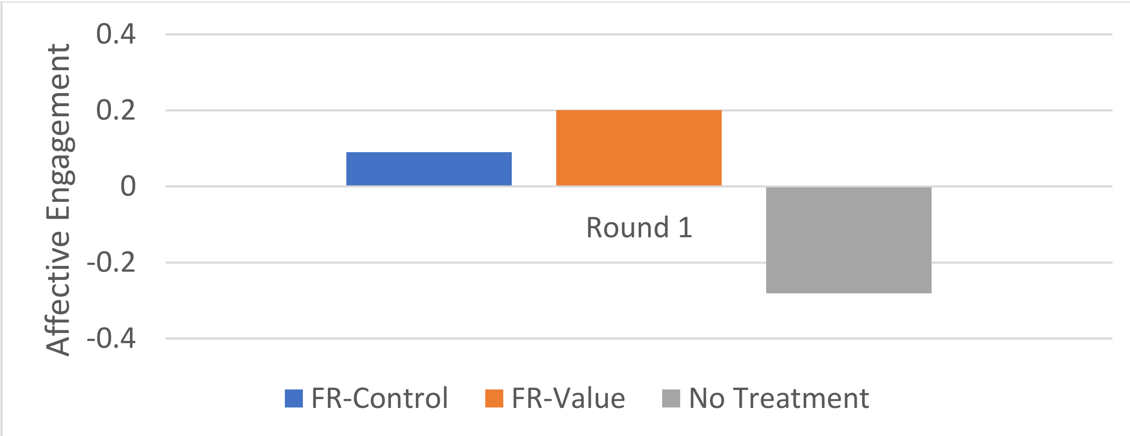


Figure 10. Affective engagement mean average z-scores for round 1.

The means and standard deviations across conditions I summarize in Table 6 below. There was a statistically significant main effect for condition for round 1, $F(2, 132) = 3.80$, $p = .025$, $\eta p^2 = .054$, but not for round 2, $F(2, 120) = 2.25$, $p = .11$, $\eta p^2 = .036$, or round 3, $F(2, 120) = 1.22$, $p = .30$, $\eta p^2 = .022$. There was no statistically significant main effect of Initial Geometry Performance for round 1, $F(2, 132) = 0.010$, $p = .75$, $\eta p^2 = .001$, or round 2, $F(2, 120) = 1.89$, $p = .17$, $\eta p^2 = .015$, or round 3, $F(2, 120) = 0.14$, $p = .71$, $\eta p^2 = .001$. There was no significant FR condition x Initial Geometry Performance interaction for round 1, $F(2, 132) = 0.54$, $p = .58$, $\eta p^2 = .008$, or round 2, $F(2, 120) = 0.38$, $p = .68$, $\eta p^2 = .006$, or round 3, $F(2, 120) = 0.57$, $p = .57$, $\eta p^2 = .009$. Post hoc testing using LSD comparisons indicated, during round 1, participants in FR-Control had significantly higher ($p = .042$) affective engagement scores than the no treatment group. Additionally, students in the FR-Value condition had significantly higher ($p = .009$) affective engagement scores than the no treatment group. The two FR groups did not significantly differ on this measure ($p = .54$).

Table 6. *Mean averages and standard deviations of affective engagement scores by condition and initial performance for three rounds.*

FR Condition	Geo Performance	Round 1		Round 2		Round 3	
		M	*SD*	*M*	*SD*	*M*	*SD*
FR-Control	High	0.19	0.96	0.18	0.92	0.14	0.87
	Low	-0.02	0.76	-0.23	0.79	-0.15	0.80
	Total	0.09	0.87	-0.02	0.86	<-0.01	0.84
FR-Value	High	0.12	1.03	0.28	0.95	0.10	1.07
	Low	0.28	0.82	0.14	0.85	0.19	0.79
	Total	0.20	0.91	0.21	0.89	0.15	0.93
No Treatment	High	-0.24	0.93	-0.14	0.86	-0.16	0.85
	Low	-0.32	0.71	-0.23	0.81	-0.13	0.74
	Total	-0.28	0.81	-0.19	0.82	-0.14	0.79
Total	High	0.02	0.97	0.11	0.91	0.03	0.93
	Low	-0.02	0.79	-0.11	0.82	-0.03	0.78
	Total	0.00	0.88	0.00	0.87	0.00	0.86

Inspection of the mean values shows that each of the FR treatments were higher than compared to its respective no treatment group. The statistically significant main effect in round 1, and the post hoc testing suggest FR can influence students' affective engagement as measured in this study, although the effects diminished across the three rounds.

Behavioral Engagement

The second research question asked whether the FR conditions differentially predicted students' behavioral engagement. I conducted a MANOVA on the main effect of the independent variables (FR conditions, initial geometry performance) on the dependent variables (disengagement scores and time-off-task). Using the FR condition (FR-Control, FR-Value, no treatment) and initial geometry performance (high and low) as the independent variable and behavioral engagement as the dependent variable, I conducted the MANOVA for rounds 1, 2, and 3, respectively. For these analyses, I predicted there would be a main effect of the FR condition compared to the no treatment group. Also, I expected students in the FR-Value who were low-performers would show more significant improvement than high-performing students. I expected the opposite for FR-Control where high-performing students were thought to see the most benefit.

A two-way MANOVA on the main effect of FR condition on disengagement and time-off-task for round 1 was statistically significant, Wilks' λ = .92, $F(2, 132) = 2.98$, $p = .020$, $\eta p^2 = .043$, and there was statistically significant main effect of Initial Geometry Performance, Wilks' λ = .93, $F(2, 132) = 5.06$, $p = .008$, $\eta p^2 = .072$. There was no significant FR condition x Initial Geometry Performance interaction in round 1 Wilks' λ = .99, $F(2, 132) = 0.16$, $p = .96$, $\eta p^2 = .003$

A two-way MANOVA on the main effect of FR-condition on disengagement and time-off-task for round 2 was not statistically significant, Wilks' λ = .95, $F(2, 120) = 1.65$, $p = .16$, $\eta p^2 = .027$, but there was a statistically significant main effect for Initial Geometry Performance, Wilks' λ = .95, $F(2, 120) = 3.37$, $p = .038$, $\eta p^2 = .054$. There was no significant FR condition x Initial Geometry Performance interaction in round 2 Wilks' λ = .99, $F(2, 120) = 0.039$, $p = .99$, $\eta p^2 = .001$.

A two-way MANOVA on the main effect of FR condition on disengagement and time-off-task for round 3 was not statistically significant, Wilks' λ = .96, $F(2, 120) = 1.20$, $p = .31$, $\eta p^2 = .020$, and there was no statistically significant main effect for Initial Geometry Performance, Wilks' λ = .96, $F(2, 120) = 1.02$, $p = .37$, $\eta p^2 = .017$. There was no significant FR condition x Initial Geometry Performance interaction in round 3 Wilks' λ = .93, $F(2, 120) = 2.06$, $p = .087$, $\eta p^2 = .034$.

Since the MANOVA uncovered statistical main effects for condition and Initial Geometry Performance in round 1 additional analyses were conducted. The means and standard deviations across conditions for these two variables in rounds 1 are summarized in Table 7. Using an ANOVA, I explored the effect of the condition, performance, and interaction on disengagement and time-off-task separately for round 1.

Table 7. *Mean averages and standard deviations of disengagement and time-off-task by condition and initial performance for round one.*

		Round 1			
FR	Geo	Disengage		Time Off	
Condition	Performance	*M*	*SD*	*M*	*SD*
FR-Control	High	5.04	1.40	3.74	0.70
	Low	5.50	1.10	3.79	0.72
	Total	5.26	1.08	3.76	0.70
FR-Value	High	5.38	1.07	3.79	0.59
	Low	5.75	1.36	3.91	0.90
	Total	5.58	1.23	3.85	0.76
No Treatment	High	5.83	1.44	4.12	0.87
	Low	6.38	1.28	4.27	0.98
	Total	6.11	1.37	4.20	0.92
Total	High	5.41	1.23	3.88	0.74
	Low	5.89	1.29	4.00	0.89
	Total	5.65	1.28	3.94	0.82

Figure 11 below graphs the dependent variable, disengagement, to show the mean differences between the three conditions for round 1. I found there was a statistically significant main effect for disengagement in the FR condition, $F(2, 132) = 5.46$, $p = .005$, $\eta p^2 = .076$, and Intial Geometry Performance $F(2, 132) = 4.80$, $p = .030$, $\eta p^2 = .035$ in round 1. There was no significant FR-condition x Initial Geometry Performance interaction $F(2, 132) = 0.062$, $p = .94$, $\eta p^2 = .001$. Post hoc testing using LSD comparisons indicated, during round 1, participants in

FR-Control had significantly lower (p = .001) disengagement scores than the no treatment group. Additionally, the FR-Value condition had significantly lower (p = .041) disengagement scores than the no treatment group. The two FR groups did not significantly differ for this measure (p = .22).

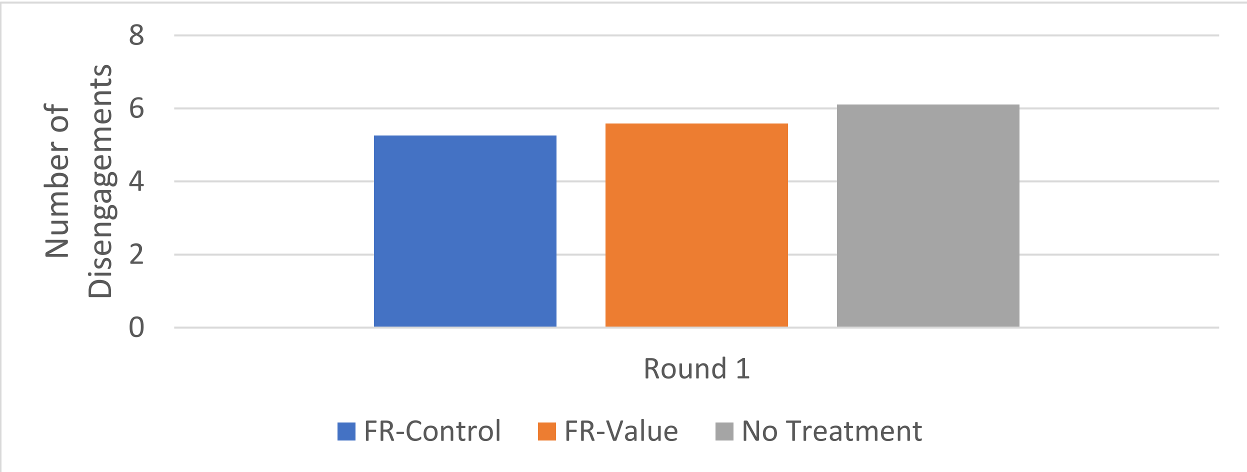


Figure 11. Disengagement mean average scores for each condition during round 1.

Figure 12 graphs the dependent variable, time-off-task, to show the mean differences between the three conditions for round 1. I found here was a statistically significant main effect for the FR-condition, $F(2, 132) = 3.62$, $p = .029$, $\eta p^2 = .052$. There was no statistically significant main effect of Initial Geometry Performance $F(2, 132) = 0.66$, $p = .42$, $\eta p^2 = .005$ found for round 1. There was no significant FR-condition x Initial Geometry Performance interaction $F(2, 132) = 0.049$, $p = .95$, $\eta p^2 = .001$. Post hoc testing using LSD comparisons indicated, during round 1, participants in FR-Control had significantly lower (p = .012) time-off-task scores than the no treatment group. Additionally, the FR-Value condition had significantly lower (p = .044) time-off-task scores than the no treatment group. The two FR groups did not significantly differ for this measure (p = .62).

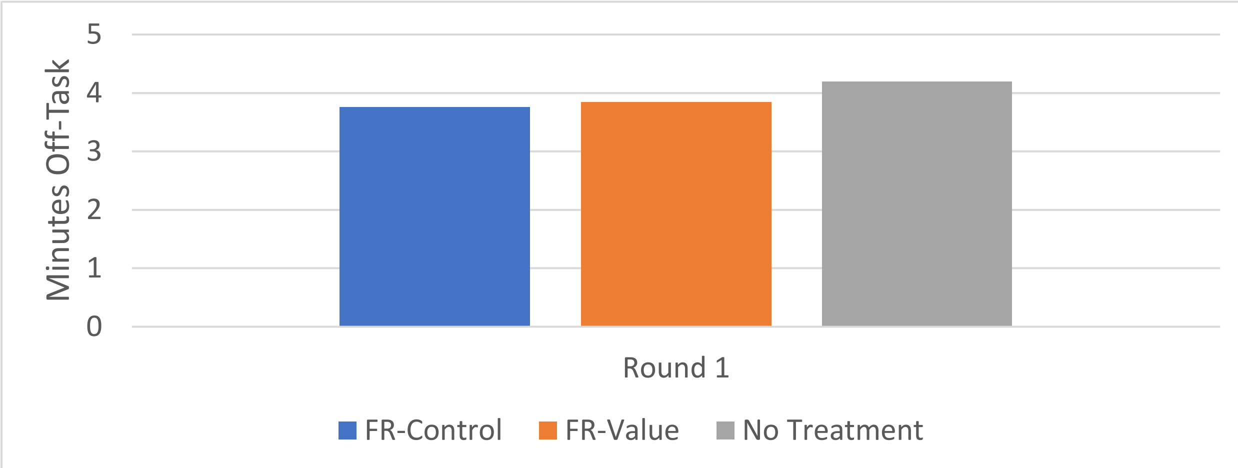


Figure 12. Time-off-task mean average times for each condition during round 1.

Inspection shows the mean averages for both FR treatments were lower than compared to their respective no treatment group (here lower indicates less disengagement or time-off-task, i.e., improved behavioral engagement). The statistically significant main effects in round 1, and the post hoc testing suggest FR can influence students' behavioral engagement as measured in this study. The individual FR treatments did not influence high- or low-performing students differently for this measure.

Learning

The third research question asked whether the FR conditions differentially predicted students' learning of the geometry concepts, as assessed by the ITS geometry post-test. I conducted a 3 x 2 repeated measures ANOVA using condition as three-level between-subjects factor (FR-Control, FR-Value, and no treatment) and initial performance as a two-level between-subjects factor (high and low). For these analyses, I expected there would be a main effect of both FR conditions compared to the no treatment group. Also, I predicted students in the FR-Value who were low-performers would have higher scores than high-performing students. I expected the opposite for FR-Control where high-performing students were thought to see the most benefit. Figure 13 graphs the dependent variable, learning of geometry, to show the mean differences between the three conditions.

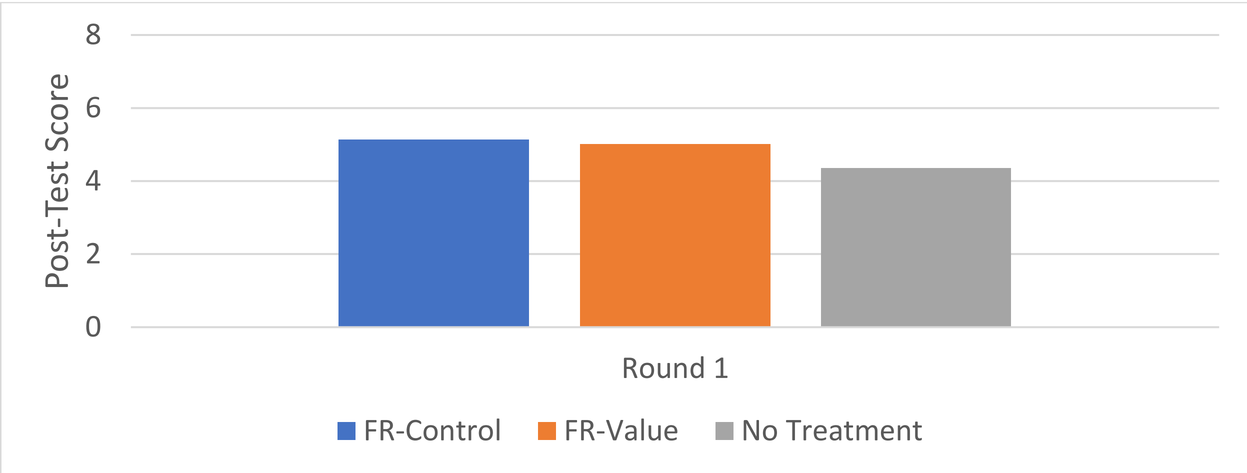


Figure 13. Mean average post-test scores for each condition.

The means and standard deviations across conditions I summarize in Table 8. I found there was a statistically significant main effect for FR-control $F(2, 120) = 3.97$, $p = .021$, $\eta p^2 = .062$, and for Initial Geometry Performance, $F(2, 120) = 16.88$, $p = <.001$, $\eta p^2 = .12$. There was no significant FR condition x Initial Geometry Performance interaction $F(2, 120) = .50$, $p = .61$, $\eta p^2 = .008$. Post hoc testing using LSD comparisons indicated participants in FR-Control had significantly higher ($p = .010$) post-test scores on the ITS Geometry Test than the no treatment group. Additionally, the FR-Value condition had significantly higher ($p = .029$) post-test scores than the no treatment group. The two FR groups did not significantly differ from each other for this measure ($p = .68$). The statistically significant main effect, and the post hoc testing suggest FR can influence students' learning of geometry as measured in this study. The individual FR treatments did not influence high- or low-performing students differently for this measure.

Table 8. *Mean averages and standard deviations of post-test scores by condition and initial performance.*

FR	Geo	Post Test	
Condition	Performance	*M*	*SD*
FR-Control	High	5.78	1.37
	Low	4.49	1.31
	Total	5.13	1.48
FR-Value	High	5.37	1.44
	Low	4.66	1.24
	Total	5.01	1.38
No Treatment	High	4.84	1.39
	Low	3.90	1.29
	Total	4.35	1.43
Total	High	5.33	1.43
	Low	4.35	1.30
	Total	4.84	1.45

*Note: The post-test values ranged from 0 to 8.

Retention. An additional consideration was how FR affected the performance of students' retention versus transfer of leaning. I conducted an ANOVA using the FR condition (FR-Control, FR-Value, and no treatment) and Initial Geometry Performance (high and low) as independent variables and retention measured from scores on questions 1, 3, 5, and 7 of the ITS geometry post-test (Appendix A) as the dependent variable.

The means and standard deviations across conditions are summarized in Table 9. There was no statistically significant main effect for FR condition $F(2, 120) = 2.46$, $p = .090$, $\eta p^2 = .039$, but there was for Initial Geometry Performance, $F(2, 120) = 12.11$, $p = .001$ $\eta p^2 = .092$. There was no significant FR condition x Initial Geometry Performance interaction, $F(2, 120) = 0.31$, $p = .73$ $\eta p^2 = .005$. While some statistically significant effects were seen the FR condition had limited influence on retention of geometry concepts as measured in this study.

Table 9. *Mean averages and standard deviations of retention scores by condition and initial performance.*

FR Condition	Geo Performance	Retention Scores *M*	*SD*
FR-Control	High	3.16	0.52
	Low	2.75	0.55
	Total	2.95	0.57
FR-Value	High	3.10	0.53
	Low	2.86	0.59
	Total	2.98	0.57
No Treatment	High	2.93	0.59
	Low	2.53	0.61
	Total	2.72	0.63
Total	High	3.06	0.55
	Low	2.71	0.59
	Total	2.87	0.59

*Note: The measure of retention ranged from 0 to 4.

Transfer of learning. To further distinguish how FR influences the transfer of the geometry concepts learned, I conducted an ANOVA using the FR condition (FR-Control, FR-Value, and no treatment) and initial geometry performance (high and low) as independent variables and transfer of learning measured from scores on questions 2, 4, 6, and 8 of the ITS geometry post-test (Appendix A) as the dependent variable.

The means and standard deviations across conditions are summarized in Table 10. There was a statistically significant main effect for FR condition $F(2, 120) = 4.50$ $p = .013$, $\eta p^2 = .070$, and for Initial Geometry Performance, $F(2, 120) = 17.06$, $p = <.001$ $\eta p^2 = .12$. There was no significant FR-condition x Initial Geometry Performance interaction, $F(2, 120) = 0.50$, $p = .61$ $\eta p^2 = .008$. Post hoc testing using LSD comparisons indicated participants in FR-Control had significantly higher ($p = .005$) transfer of geometry learning scores than the no treatment group.

Additionally, the FR-Value condition had significantly higher (p = .035) transfer-of-learning scores than the no treatment group. The two FR groups did not significantly differ for this measure (p = .45). The statistically significant main effect, the post hoc testing suggest FR can influence students' transfer of geometry concepts as measured in this study.

Table 10. *Mean averages and standard deviations of transfer-of-learning scores by condition and initial performance.*

FR Condition	Geo Performance	TOL Scores *M*	TOL Scores *SD*
FR-Control	High	2.62	0.88
	Low	1.74	0.82
	Total	2.18	0.95
FR-Value	High	2.27	0.97
	Low	1.80	0.73
	Total	2.04	0.88
No Treatment	High	1.90	0.92
	Low	1.37	0.81
	Total	1.64	0.92
Total	High	2.27	0.96
	Low	1.63	0.80
	Total	1.95	0.93

*Note: The measure of transfer-of-learning ranged from 0 to 4.

CHAPTER 6:

DISCUSSION

The primary purpose of this study was to conduct an experimental test of the effectiveness of using formative reappraisal to increase affective engagement, behavioral engagement, and enhancing learning. The subsequent discussion focuses on the broad pattern of the study's findings, before addressing the implications, limitations, and future directions, and ending with concluding remarks.

Overview of Major Findings

The results support four main findings relating to the use of formative reappraisal (FR): a) Influence of FR on perceptions of value and control, b) student affective and behavioral engagement improvements, c) student learning, and d) interaction between high- and low-performing students. I cover these in turn next.

Influence of FR on Perceptions of Value and Control

The first result pertained to the manipulation checks of the formative reappraisal intervention. For every round of the study, results on the motivation questionnaire showed statistically significant differences for the respective FR group. Students in the FR-Value condition reported higher levels of perceived task value in the subject while students in the FR-Control reported a significant change in the perceived competence of the topic. This finding was important because verifying FR did influence the appropriate student perceptions allowed me to give less weight to the alternative explanations for further findings.

Student Affective and Behavioral Engagement Improvements

The second important result was that students in both FR conditions showed significantly decreased observable disengagement with significantly increased self-reported affective engagement in round 1. The concerns raised in the literature review about students easily becoming disengaged without human interaction highlight the importance of this finding. In this context, students did not ignore the FR prompts but chose to engage in them routinely.

In general, their responses showed students took the request to engage in reappraisal quite seriously and often left thoughtful replies. Representative responses included:

- Geometry is exciting because it allows us to interact with our world.
- Advisers need to inform first-year students how useful geometry will be to their lives.
- This geometry problem is complex, and still, there’s a hidden beauty in it.
- Geometry is a fun course that can help students professionally and personally.

Only rarely did students give what appeared to be a meaningless response or no response at all.

The nature of the study allowed for a wide variety of possible response patterns in the students’ reaction to a just-in-time reappraisal. It was conceivable learners would find the intervention intrusive rather than positive. Repeatedly asking questions about the value of the topic or how to teach the current problem could have backfired causing students to zone out or ignore requests. Typical reactions from the students during the study suggest this was not the case.

The study indicated no significant effects for engagement in rounds 2 and 3. There were, however, lower levels of overall observed disengagement and higher levels of self-reported affective engagement for the FR groups compared to the no treatment group in each round. Although, the effect size diminished in the conditions for each successive round. A plausible reason is discussed in the limitation section. Additional testing in the future is recommended to explore these initial findings using a similar, multi-round treatment design as many educational leaders agree that learning gains need sustained engagement (Loveless, 2015; PISA, 2012).

Student Learning

The third important result was that FR positively influenced learning of geometry, as measured in the study. Both FR groups saw an improvement on the post-test and both had modest effect sizes. Representation of meaningful learning was measured by the transfer of learning questions.

Applying these questions from the post-test, I found significant improvements for students in both FR conditions. As these were some of the largest effect sizes revealed by both FR conditions in the study, the results suggested a potential advantage to the FR approach.

Interaction Between High- and Low-Performance Students

Fourth, there was no significant interaction found between the FR conditions and Initial Geometry Performance in any of the statistical testing. A major hypothesis of the study was that the reappraisal type performed (control or value) would influence the type of student (high- or low-performance) differently. That hypothesis found no support in the study. Possible reasons for the lack of interaction between FR condition and Initial Geometry Performance and suggestions for further research I describe in the *Recommendations* section below.

Implications

The current study had both theoretical and applied implications. I address them separately.

Theoretical Implications

The study identified two theoretical implications: a) The findings supported the control-value theory model and b) the results bolstered the notion of appraisals being treated as a type of schema. I discuss both in this section.

First, the results of this study bolster the connections and mediating mechanisms submitted by Pekrun's (2006) control-value theory. Pekrun explained this change in appraisal ultimately influences the student's academic achievement. This study showed comparable results, where students who used cognitive reappraisal to address control and value issues demonstrated better results for engagement and learning. An analysis of the motivation questionnaire showed students in each respective FR group rated appraisals of the task differently from the group that did not use FR. The control-value theory does not differentiate cognitive appraisals based on high- or low-performing students. The lack of interaction found in this study between FR types and student performance types is therefore still consistent with Pekrun's theory. The influence of FR on the precedent achievement emotion is an area that would benefit from future study.

Second, the results showed general agreement with the positive effects that emotion regulation can provide students learning in a computer-based environment (Azevedo & Cromley, 2004; Azevedo et al., 2004b). Examining emotion regulation literature further, however, highlighted an unexpected study finding related to the FR timing.

The consensual process model breaks emotion regulation into two methods: antecedent-focused and response-focused (Gross, 1998). Theorists suggest restructuring a schematic network requires substantial effort across time (Taber, 2010) and therefore changing appraisal schemas would require many trials. Following the reasoning that appraisals are done in these schematic networks, I used both the antecedent-focused and response-focused strategies in this study.

The positive results found from formative reappraisal, especially regarding learning, ran counter to a large body of research in the field. Investigations indicate antecedent-focused approaches yield many positive benefits whereas response-focused strategies produce many deleterious effects (Higgins, 1987; John & Gross, 2004; Sheldon, Ryan, Rawsthorne, & Ilardi, 1997). The response-focused method has been shown to be particularly problematic for student cognitive processing and memory function (Dillon et al., 2007; Hayes et al., 2010; Richards & Gross, 1999, 2000; Sheppes & Meiran, 2007, 2008).

Gross (1998) posits the dilemma of different effects is a matter of timing. If the goal is upending an emotion, then the antecedent-focused strategy comes early in the process while the response-focused approach comes late in the process. Researchers suggest by reacting late after the behavior has already occurred requires more effort for students to manage the emotional responses, and repeated attempts exhaust cognitive reserves (Higgins, 1987; Gross, 1998). There is broad agreement on this timing problem (Hayes et al., 2010; Higgins, 1987; John & Gross, 2004; Sheldon et al., 1997).

The FR intervention in this study used both an antecedent-focused and response-focused approach, yet I found no indication of decreased learning resulting from this process in the study. Results revealed the opposite—post-test scores significantly increased in the FR conditions, and the effect was stronger for learning transfer, which I suggest required deeper cognitive processing. Reviews of the response-focused approach mainly concentrate on the same emotion regulation, expressive suppression (Dillon et al., 2007; Hayes et al., 2010; Higgins, 1987; John & Gross, 2004; Richards & Gross, 1999, 2000; Sheldon et al., 1997; Sheppes & Meiran, 2007, 2008). While inconclusive, the study results raise the possibility the method used in emotion regulation is a more substantial consideration than the timing of when learners use it. Since this study used a combination of approaches, I drew no conclusion, and further investigation is needed, as I discuss in the recommendations section below.

Applied Implications

Educational technology is transforming how students learn about the world. As learners immerse themselves, it holds exciting potential. At the same time, the sea of information and the disconnect from human interaction during the process opens the door to many potential affective states. The research demonstrated formative reappraisal (FR) is one viable option for supporting students using an ITS. Once the emotions or behaviors of the end-user are detected, implementing FR is a low-cost solution. Developers could integrate FR into most computer-based environments as it only requires a simple script to prompt students. As students increasingly turn towards nontraditional learning environments, FR offers a promising method to assist in regulating their emotions to improve the learning experience.

As technology improves, so does the variety of techniques to discern a user's affect in ITSs (Calvo & D'Mello, 2010). As emotion detection increases, integrated strategies that consider both cognition and emotion to interact efficiently and productively with learners are needed. Formative reappraisal has potential to help learners manage the emotional experiences that inevitably arise during the learning process. The end goal of FR is to encourage users to self-regulate. The formative approach laid out in this paper used just-in-time reminders, which researchers suggest helps students develop habits (Nahadi et al., 2015). After a certain number of training sessions, it may not be required to continually prompt learners. Systems may be built to help students be more aware of their emotional state allowing them to decide when it is necessary to use strategies

Limitations

This study, like most, has limitations related to validity and reliability. Individually, I consider six limitations. Each of the three dependent variables (affective engagement, behavioral engagement, and learning) plus three additional considerations had a combination of validity and reliability issues that I discuss below.

First, the affective engagement measure introduced internal validity concerns stemming from student self-reporting and when they were reported. Scores were assumed to be an indication of the learner's affective engagement with geometry content in the ITS module. What determined the genesis of the affective component, however, is uncertain. Because I obtained these scores in proximity to the students finishing the module, it was assumed their assessments of feelings concerned the learning process, but other issues could be reflected. For example, students could have been reporting they were happy the session was over, annoyed they needed to complete additional surveys, or frustrated at having to schedule time for more modules. None of these states, as reported, would reflect success or failure of the FR strategy. Also, since students in the FR groups completed different tasks than the no treatment group it is possible they considered themselves more engaged because of the different activities, and not because they were more engaged with the subject matter.

Second, the behavioral engagement variable used the proxy measures of time-off-task and number of observed disengagements and not a direct measure. What appeared to be the learner disengaged could have been a student thinking deeply about the problem. For instance, if a learner mentally rotates a geometric figure, it takes considerable cognitive resources. It is conceivable that doing this would result in looking away from the screen thus appearing to be distracted. If what appeared to be learners losing focus was instead students thinking more in-depth about the topic, it is possible this action would manifest itself in the transfer of learning section of the post-test, which required more than rote memorization of the material. Participants in the no treatment group, however, on average, looked away from their screens more often, and had significantly lower scores on transfer of learning questions. The combined findings weaken the suggestion that disengagements resulted from students thinking deeper about the subject.

Additionally, the time-off-task variable has a potential confounding variable. Only the FR groups received a separate task when they became disengaged, which redirected them back to the module. Since an outside observer interrupted the disengagement, it is possible that introducing a novel task that ultimately steered students back to the problem is what reduced the time-off-task and not the reappraisal. The fact that the behavioral engagement analysis first examined for a joint effect using both disengagement and time-off-task reduces some concerns. Nevertheless, researchers could address limitations for these measures in future studies.

A solution for improving measuring disengagement could be activating a prompt on the system that asks the student if they considered themselves disengaged each time disengagement is counted. This approach, however, could be overly invasive, and it is possible students would be reluctant to admit disengagement occurred. Another option would be to capture video of the recorded instances of disengagement. Afterward, students could observe the video of the supposed disengagement and confirm or deny the event. The post hoc approach would allow a researcher to develop rater reliability with the participant feedback. Both methods could improve potential validity and reliability issues by using more than experimenters' observations and including student input. To improve the time-off-task measure, the no treatment group could also complete a novel task that takes approximately the same amount of time as the reappraisal.

Third, I measured the learning variable with a self-created post-test. Half of the post-test questions I generated from material covered in the ITS modules. These questions were simple variants of what students were previously asked and required only rote memorization to answer correctly. The analysis showed no significant main effect between groups for these questions. It is possible the problems were not sufficiently challenging.

Fourth, the geometry modules likely got considerably harder as they progressed was a limitation. The final modules had students constructing full geometry proofs. The motivation questionnaire supports the idea of problems getting harder. The perceived competence scores for all groups declined with each round and the difference between the groups also shrunk. Since the content was not similar enough across rounds, it is difficult to determine trends in the data. For example, for the perceived competence questions the effect for the FR-Control got smaller with each round. It is possible as the content increased in difficulty it influenced students affective and behavioral engagement levels, which declined in each successive round. Was this because the FR strategy was wearing off, or because the content increased in difficulty?

Fifth, the purpose of the FR-Control condition was to make the task more cognitively taxing for students by requiring them to imagine how they would teach the geometry problem. The manipulation check showed that students in this condition had significantly lower perceived competence scores. At the same time, however, the challenge had a potential dual benefit of making students think more in-depth about the problem. This secondary benefit introduces a possible confounding variable where students could achieve positive results because of the cognitive scaffold and not the reappraisal. An additional experiment that compares a scaffold, a reappraisal, and combination of both would provide more clarification.

Sixth, the duration of study only represented approximately one week of what is typically a 15-week semester. The length of time students worked in the study is a threat to external validity. It is possible this was not enough time for the study to induce a larger effect or, conversely, it is possible the effects I did discover may typically fade over the course of a regular semester.

Recommendations

The results of using formative reappraisal suggest additional studies. Understanding the timing issues of FR and how this relates to student performance improvement is needed. A future study examining the three conditions of cognitive reappraisal, formative reappraisal, and a combination of both (as this study used) could provide important insights into the importance of the emotion regulation timing.

Additionally, an important consideration is the lack of interaction between the conditions and the Initial Geometry Performance groups. The FR condition, whether control or value, worked well across conditions. Applying the control-value theory model does not lead to the conclusion that high- and low-performing students must apply different cognitive appraisals. For example, a low-performing student may highly value a task even though he is not good at executing it. Research does show high- and low-performing students tend to experience different achievement emotions (Gumora & Arsenio, 2002; Pekrun et al., 2011), and an assumption was made this would lead to different cognitive appraisals.

Future work should consider a more nuanced investigation into the affective states experienced by the participants. Researchers using the control-value theory model would likely benefit more from knowing the student achievement emotion before and after the FR intervention than knowing if the student is high- or low-performing. Recent advances have shown promising accuracy in detecting students' affect from facial expressions and gross body movements using computer vision and machine-learning techniques (Bosch, D'Mello, Ocumpaugh, Baker, & Shute, 2016). These latest advancements could provide new avenues for research, which I discuss next.

Future Directions

Currently, it is uncommon that a commercially built ITS utilizes a disengagement detector. Those systems with such sensors are mainly for research purposes, and the focus of the resulting interventions has been primarily on student learning strategies (Strain & D'Mello, 2015). New technological advances are increasing available options. For example, a breakthrough is occurring with cameras—they are getting brains. The new generation of cameras no longer only sees what is put in front of them but can act on it (Manjoo, 2018). This development is creating intriguing possibilities. It does not take long to imagine how educational researchers can use cameras that can interpret the world.

The new generation cameras increase the feasibility of intelligent systems that can not only detect disengagement but other affective states as well. This technology is rapidly developing, and it holds promise for ITSs because a successful reengagement strategy may require recognizing causes of disengagement. For example, students who become disengaged because they are bored may need entirely different remedies than students who lose focus because of frustration (Baker, D'Mello, Rodrigo, & Graesser, 2010). Improved camera technology holds the key to better discrimination for ITSs using adaptive automation.

Developers have focused on accessing finer levels of a student's cognition to improve the instruction (the granularity hypothesis) while ignoring the emotional and affective components of learning. This scheme worked well when the system mainly had access to student keystrokes and mouse clicks. Identifying emotional aspects often depends on a visual component (e.g., body posture, facial expressions). With the improved camera technology, the next generation of ITSs can be built with a new assumption that finer granularity of learners' emotional states can be beneficial. Distinguishing not only between different affective states but increasing the granularity to see differences within the same state could prove insightful. Imagine a system that can regularly identify bored students. The detection would be helpful, yet the control-value theory suggests multiple reasons why boredom could occur such as working with tasks that are too complicated, too simplistic, uninteresting, or lack value (Pekrun et al., 2010). Interventions designed to re-engage students will likely require more than only identifying the affective state but understanding the reason it occurred in the first place.

Systems like IBM's Watson, which is slated to move into higher education classrooms this year, hold the most significant promise (IBM Pearson, 2016). These systems can gather enormous amounts of data on a global level. Watson integrates massively parallel processors and is built on IBM's DeepQA technology, which is used to generate hypotheses, gather large bodies of evidence, and analyze data (Ferrucci et al., 2010). If smart cameras can readily recognize student affect, then parallel processing systems can capture the data and identify trends. For instance, it would not be surprising for Watson to discover bored students who understand the material respond better to certain interventions than bored students who are confused. By combining data from diverse sources around the globe, these systems can uncover insights in seconds, which may have taken investigators years of educational research.

Concluding Remarks

Upon completion of this study, there was an announcement that Cedar Valley College administrators and faculty decided to turn over the primary instruction of the entire developmental mathematics program to the intelligent tutor system, ALEKS. One-hundred percent of the faculty supported the decision after witnessing first-hand how beneficial the system has been for students. Indeed, more mathematics classes will likely follow, and some chemistry courses at the college have started using these systems as a supplement. Similar occurrences are happening in school systems throughout the country and once word spreads of the result there could be a cascading effect (Sparks, 2017).

As the use of these systems continues to expand it is more important than ever to consider other aspects of the human-computer interaction. This study extends the existing research on this interaction by providing new empirical evidence for using emotion regulation to improve engagement and learning while using an ITS. Research concerned with connecting ITS design to an integrative cognitive and emotion theory is still in its infancy. Although, during the last decade, there has been a substantial increase in educational research concerning the cause, effect, and interventions of disengagement in learning.

The preliminary findings using formative reappraisal are a reason for optimism. Adverse effects typically associated with response-focused emotion regulation strategies were not evident with the FR approach. As learners interact with more technology, the FR approach offers a promising method for promoting participation. Developing effective interventions that encourage engagement and support an enjoyable learning experience is the path forward.

APPENDICES

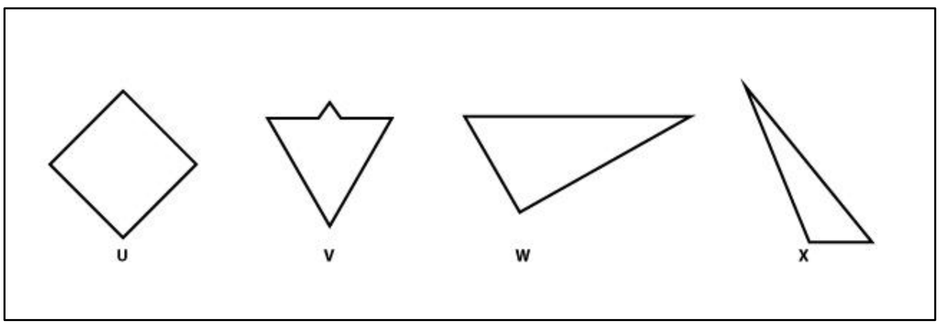
APPENDIX A — Instruments

Van Hiele Test

The van Hiele theory posits that students move through a hierarchy of stages in geometric understanding. These stages are sequential levels that a student progresses through with each level becoming increasingly sophisticated in the amount of description, analysis, abstraction, and proof. This test will cover the four levels of geometric understanding in order. Once a student has missed two questions in a row the test will stop, and a score will be recorded for the researcher to use in randomly assigning the student to a group (Usiskin, 1982).

Example of Level One Questions

Which of these are triangles?



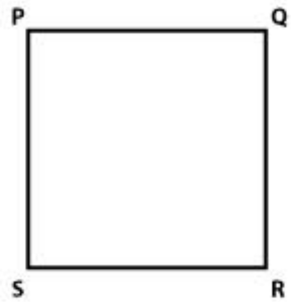
None of these are triangles.

(A) V only
(B) W only
(C) W and X only
(D) V and W only

Example of Level Two Questions

PQRS is a square.

Which relationship is true in all squares?



(A) $\overline{PR}$ and $\overline{RS}$ have the same length.
(B) $\overline{QS}$ and $\overline{PR}$ are perpendicular.
(C) $\overline{PS}$ and $\overline{QR}$ are perpendicular.
(D) $\overline{PS}$ and $\overline{QS}$ have the same length.
(E) Angle Q is larger than angle R.

Example of Level Three Questions

Statement S: ΔABC has three sides of the same length.

Statement T: In ΔABC, $\angle$B and $\angle$C have the same measure.

Which is correct?

(A) Statements s and T cannot both be true.
(B) If S is true, then T is true.
(C) If T is true, then S is true.
(D) If S is false, then T is false.
(E) None of (A)-(D) is correct.

Example of Level Four Questions

Here are three properties of a figure.

Property D: It has diagonals of equal length.

Property S: It is a square.

Property R: It is a rectangle.

Which is true?

(A) D implies S which implies R.
(B) D implies R which implies S.
(C) S implies R which implies D.
(D) R implies D which implies S.
(E) R implies S which implies D.

Geometry Placement Test

Please answer the questions as best you can. If you do not know the answer, it is okay to leave it blank.

1. Which of the following sets of numbers cannot represent the sides of a triangle?
(A) 9, 40, 41
(B) 7, 7, 3
(C) 3, 5, 1
(D) 7, 7, 7
(E) 9, 12, 15

2. The measure of an angle is 5 times as great as the measure of its complement. Find the measure of the angle.

3. Write the equation of the perpendicular bisector of the segment that joins the points *A* (3, -7) and *B*(5, 1).

4. Two parallel lines are cut by a transversal. The bisector of a pair of interior angles on the same side of the transversal intersect to form an angle that is
(A) always acute
(B) always right
(C) always obtuse
(D) either acute or obtuse, but never right
(E) None of the above can happen

5. Quadrilateral *ABCD* ~ quadrilateral *RSTW*. The lengths of sides of quadrilateral *ABCD* are 3, 6, 9, and 15. If the longest side of the quadrilateral *RSTW* is 20, find the perimeter of *RSTW*.

Motivation Survey

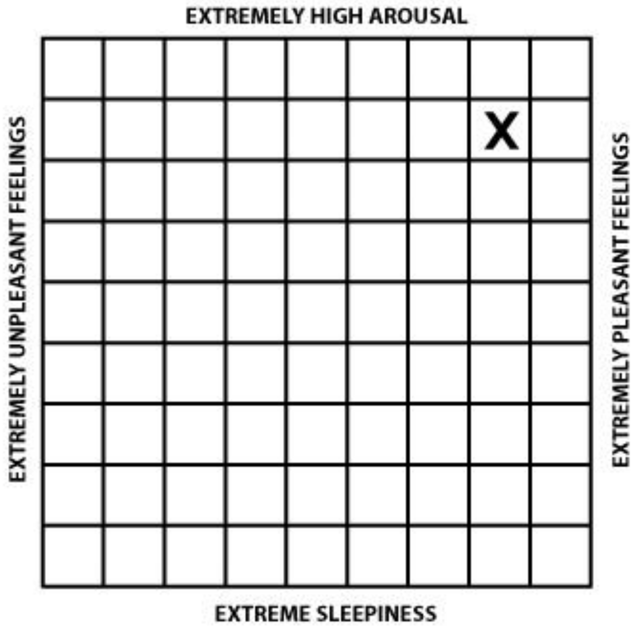
Task Value (Pintrich et al., 1991)

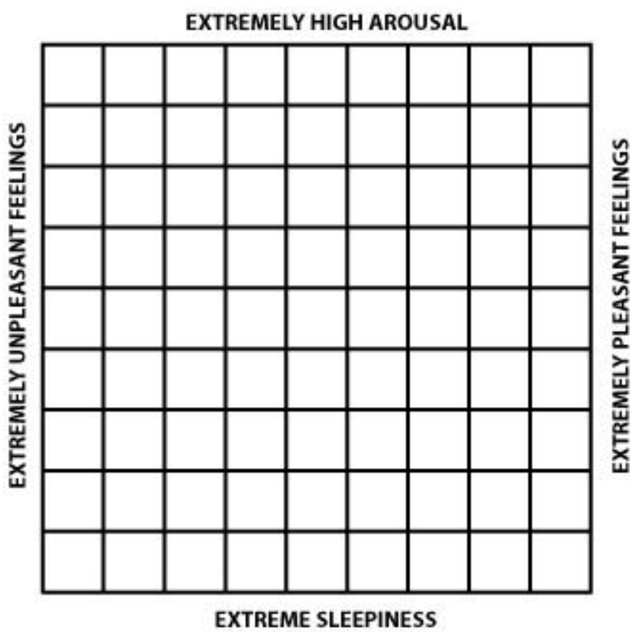
1. I think I will be able to use what I learn in this training in other areas.
3. It was important for me to learn the material that was in this training.
5. I was very interested in the content area of this training.
7. I think the material in this training was useful for me to learn.
9. I liked the subject matter of this training.
11. Understanding the subject matter of this training was very important to me.

Perceived Competence (Midgley et al., 2000)

2. I was certain I could figure out how to do the most difficult math in this training.
4. Even if the math work in this training was hard, I knew I could learn it.
6. I could easily do harder math problems than were covered if I tried.
8. I was certain I could master the skills taught in this training.
10. I could do almost all the math problems in this training if I didn't give up.

Affect Grid





The 'affect grid' is used to describe feelings. The center most point on the grid denotes neutral, average feelings which are neither positive nor negative.

The right half of the grid represents pleasant feelings while the left half represents unpleasant feelings.

The vertical dimension of the grid represents the degree of arousal. Arousal relates being wide awake or alert independent of whether the feelings are positive or nagative. The top half of the grid denotes an above average state of arousal while the bottom half represents a below average state of arousal.

For example, suppose you were suprised. The 'X' placed in the box at the top left conveys that the suprise was fairly pleasant and initiated a feeling of moderate arousal.

Please use the box on the bottom left to illustate your feelings.

(Russell et al., 1989)

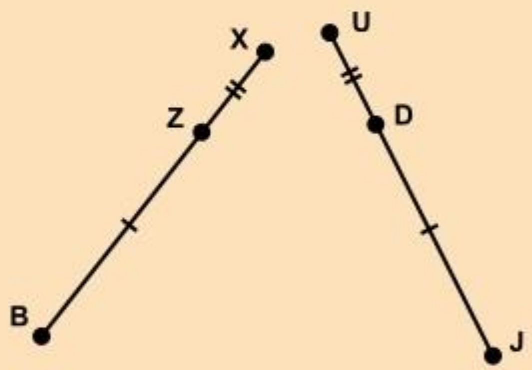
Post-Test: Types of Learning

1a) Retention question (recall)

Given:

- $\overline{XZ} \cong \overline{UD}$.
- $\overline{BZ} \cong \overline{JD}$.

Prove: $\overline{BX} \cong \overline{JU}$.



Arrange and connect the logical steps of the flow chart proof to prove $\overline{BX} \cong \overline{JU}$.

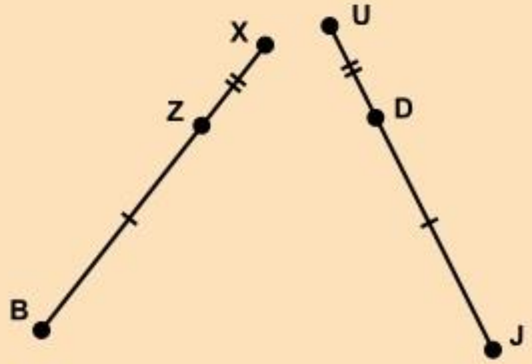
Statement	Reason
1) $\overline{XY} \cong \overline{UD}$	Given
2) $\overline{BZ} \cong \overline{JD}$	Given
3)	Segment Addition Postulate
4)	
5) $m\overline{XZ} = m\overline{UD}$	
6)	
7)	Substitution Property
8)	
9)	
10) $\overline{BX} \cong \overline{JU}$	

1b) Retention answer (recall)

Given:

- $\overline{XZ} \cong \overline{UD}$.
- $\overline{BZ} \cong \overline{JD}$.

Prove: $\overline{BX} \cong \overline{JU}$.



Arrange and connect the logical steps of the flow chart proof to prove $\overline{BX} \cong \overline{JU}$.

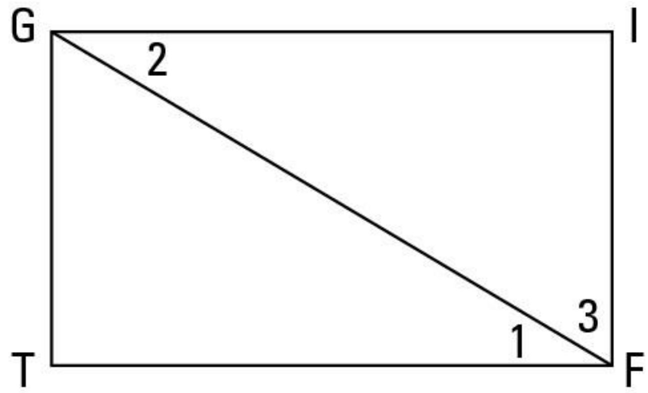
Statement	Reason
1) $\overline{XY} \cong \overline{UD}$	Given
2) $\overline{BZ} \cong \overline{JD}$	Given
3) $m\overline{BZ} + m\overline{XZ} = m\overline{BX}$	Segment Addition Postulate
4) $m\overline{JD} + m\overline{UD} = m\overline{JU}$	Segment Addition Postulate
5) $m\overline{XZ} = m\overline{UD}$	Definition of congruent segments
6) $m\overline{BZ} = m\overline{JD}$	Definition of congruent segments
7) $m\overline{BZ} + m\overline{UD} = m\overline{BX}$	Substitution Property
8) $m\overline{BZ} + m\overline{UD} = m\overline{JU}$	Substitution Property
9) $m\overline{BX} = m\overline{JU}$	Transitive Property
10) $\overline{BX} \cong \overline{JU}$	Definition of congruent segments

2a) Transfer of Learning question (near)

Given: $\angle TFI$ is a right angle

$\angle 1 \cong \angle 2$

Prove: $\angle 2$ is complementary to $\angle 3$



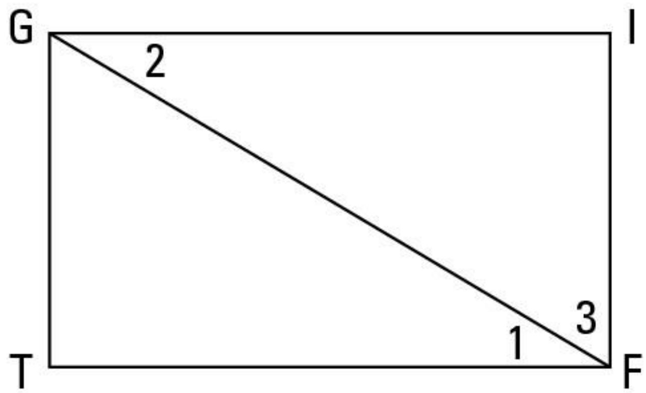
Statement	**Reason**
$\angle TFI$ is a right angle	Given
$\angle 1$ is a complementary to $\angle 3$	
$\angle 1$ is congruent to $\angle 2$	Given
$\angle 2$ is complementary to $\angle 3$	

2b) Transfer of Learning answer (near)

Given: $\angle TFI$ is a right angle

$\angle 1 \cong \angle 2$

Prove: $\angle 2$ is complementary to $\angle 3$



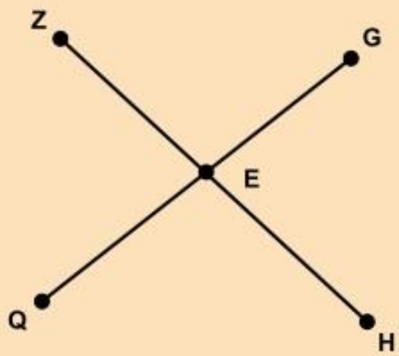
Statement	**Reason**
$\angle TFI$ is a right angle	Given
$\angle 1$ is a complementary to $\angle 3$	Definition of complementary
$\angle 1$ is congruent to $\angle 2$	Given
$\angle 2$ is complementary to $\angle 3$	Substitution property

(The Transitive and Substitution Properties, 2017).

3a) Retention question (recognition)

Given: $\overline{GQ}$ and $\overline{HZ}$ intersect at point E.

Prove: $\angle GEH \cong \angle ZEQ$.

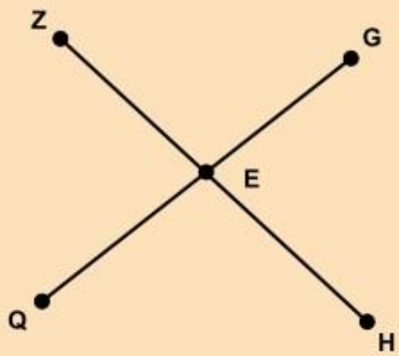


Statement	Reason
1) $\overline{GQ}$ and $\overline{HZ}$ intersect at point E	Given
2) $\angle GEH$ and $\angle HEQ$ are supplementary	a) Linear Pair Postulate, b) Definition of straight lines, c) Definition of supplementary angles, or d) Transitive Property
3) $\angle ZEQ$ and $\angle HEQ$ are supplementary	a) Linear Pair Postulate, b) Definition of straight lines, c) Definition of supplementary angles, or d) Transitive Property
4) $m\angle GEH + m\angle HEQ = 180°$	a) Linear Pair Postulate, b) Definition of straight lines, c) Definition of supplementary angles, or d) Transitive Property
5) $m\angle ZEQ + m\angle HEQ = 180°$	a) Linear Pair Postulate, b) Definition of straight lines, c) Definition of supplementary angles, or d) Transitive Property
6) $m\angle GEH + m\angle HEQ = m\angle ZEQ + m\angle HEQ$	a) Linear Pair Postulate, b) Definition of straight lines, c) Definition of supplementary angles, or d) Transitive Property
7) $m\angle GEH = m\angle ZEQ$	a) Subtraction Property of Equality, b) Definition of supplementary angles, c) Definition of congruent angles, or d) Transitive Property
8) $m\angle GEH \cong m\angle ZEQ$	a) Subtraction Property of Equality, b) Definition of supplementary angles, c) Definition of congruent angles, or d) Transitive Property

3b) Retention answer (recognition)

Given: $\overline{GQ}$ and $\overline{HZ}$ intersect at point E.

Prove: $\angle GEH \cong \angle ZEQ$.



Statement	**Reason**
$\overline{GQ}$ and $\overline{HZ}$ intersect at point E	Given
$\angle GEH$ and $\angle HEQ$ are supplementary	a) Linear Pair Postulate
$\angle ZEQ$ and $\angle HEQ$ are supplementary	a) Linear Pair Postulate
$m\angle GEH + m\angle HEQ = 180°$	c) Definition of supplementary angles
$m\angle ZEQ + m\angle HEQ = 180°$	c) Definition of supplementary angles
$m\angle GEH + m\angle HEQ = m\angle ZEQ + m\angle HEQ$	d) Transitive Property
$m\angle GEH = m\angle ZEQ$	a) Subtraction Property of Equality
$m\angle GEH \cong m\angle ZEQ$	c) Definition of congruent angles

4a) Transfer of Learning question (far)

Given:

$g = 2h$
$g + h = k$
$k = m$

Prove: $m = 3h$

Statement	**Reason**
$g = 2h$	Given
$g + h = k$	Given
$k = m$	Given
	Substitution
$3h = k$	
$3h = m$	
$m = 3h$	

4a) Transfer of Learning question (far)

Given:

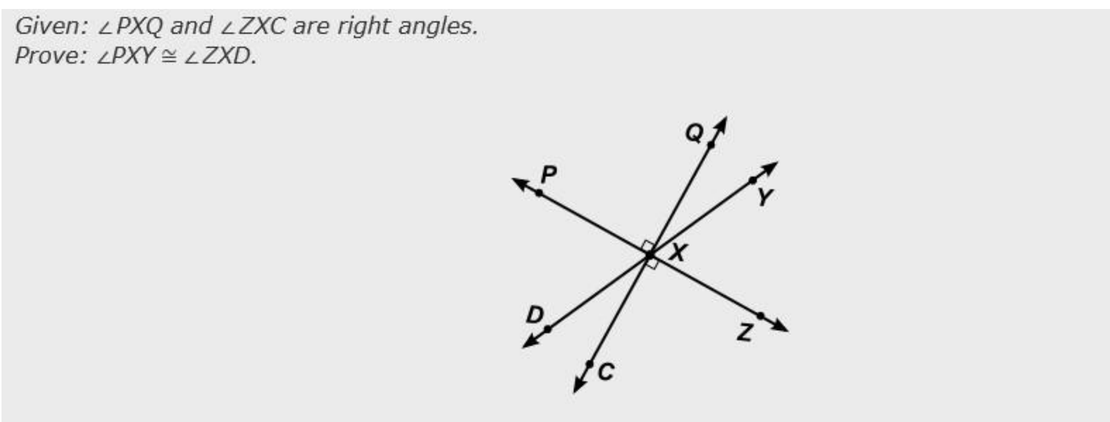
$g = 2h$
$g + h = k$
$k = m$

Prove: $m = 3h$

Statement	Reason
$g = 2h$	Given
$g + h = k$	Given
$k = m$	Given
$2h + h = k$	Substitution
$3h = k$	Simplify
$3h = m$	Transitive
$m = 3h$	Symmetric Property

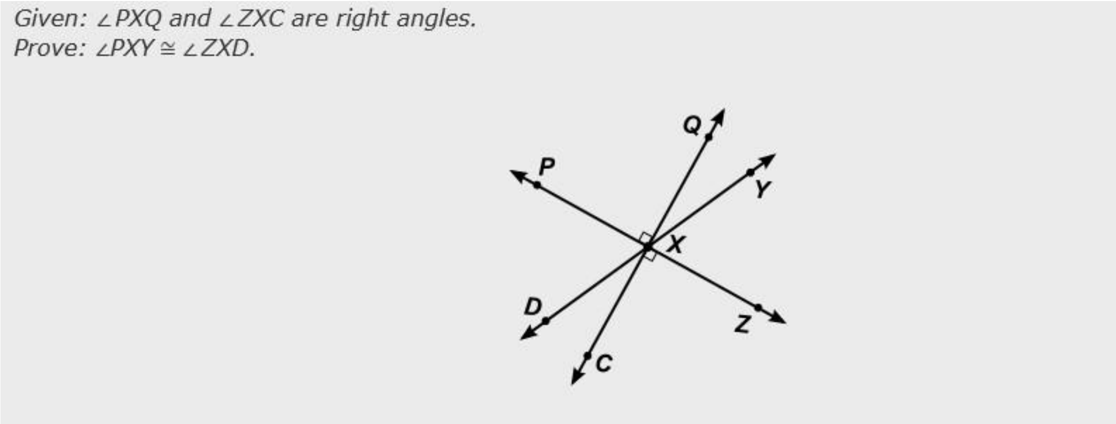
(Introducing Geometry Proofs, 2017)

5a) Retention question (recall)



Statement	Reason
∠*PXQ* is a right angle	Given
∠*ZXC* is a right angle	Given
∠PXQ ≅ ∠ZXC	
	Definition of adjacent angles
	Definition of adjacent angles
∠QXY and ∠CXD are vertical angles	
∠QXY ≅ ∠CXD	
∠PXY ≅ ∠ZXD	

5b) Retention answer (recall)



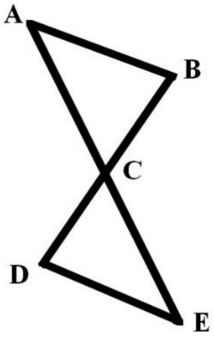
Statement	Reason
$\angle PXQ$ is a right angle	Given
$\angle ZXC$ is a right angle	Given
∠PXQ ≅ ∠ZXC	Right Angle Congruence Theorem
∠PXQ and ∠QXY are adjacent angles	Definition of adjacent angles
∠ZXC and ∠CXD are adjacent angles	Definition of adjacent angles
∠QXY and ∠CXD are vertical angles	Definition of vertical angles
∠QXY ≅ ∠CXD	Vertical Angle Theorem
∠PXY ≅ ∠ZXD	Congruent Angle Addition Theorem

6a) Transfer of Learning question (near)

If two angles of one triangle are congruent to two angles of another triangle, then the triangles are similar. This is called the Angle-Angle (AA) Theorem. When triangles are similar, it is denoted with this symbol: ~

Given: $\overline{CB} \perp \overline{BA}$,
$\overline{CD} \perp \overline{DE}$.

Prove: Δ*ABC* ~ Δ*EDC*



Statement	Reason
$\overline{CB} \perp \overline{BA}$, and $\overline{CD} \perp \overline{DE}$	Given
$\angle ABC$ and $\angle EDC$ are right angles	
∠ABC ≅ ∠EDC	
∠ACB ≅ ∠ECD	
Δ*ABC* ~ Δ*EDC*	

6b) Transfer of Learning answer (near)

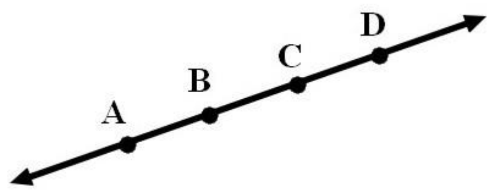
If two angles of one triangle are congruent to two angles of another triangle, then the triangles are similar. This is called the Angle-Angle (AA) Theorem. When triangles are similar, it is denoted with this symbol: ~
Given: $\overline{CB} \perp \overline{BA}$,
$\overline{CD} \perp \overline{DE}$.
Prove: $\Delta ABC \sim \Delta EDC$

Statement	Reason
$\overline{CB} \perp \overline{BA}$, and $\overline{CD} \perp \overline{DE}$	Given
$\angle ABC$ and $\angle EDC$ are right angles	Perpendicular lines intersect to form right angles
∠ABC ≅ ∠EDC	Right Angle Congruence Theorem
∠ACB ≅ ∠ECD	Vertical Angle Theorem
$\Delta ABC \sim \Delta EDC$	AA Theorem

(Leff, 1997).

7a) Retention question (recognition)

Given: AC = BD
Prove: AB = CD

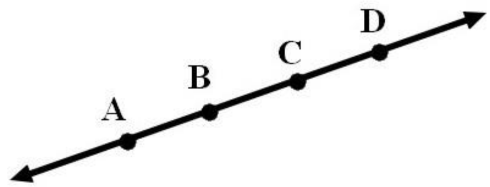


Statement	Reason
a) Subtraction Property of Inequality b) Substitution c) AC = BD d) BC = CD	Given
AB + BC = AC	a) Subtraction Property of Inequality b) Substitution c) AC = BD d) Segment Addition Postulate
BC + CD = BD	a) Subtraction Property of Inequality b) Substitution c) AC = BD d) Segment Addition Postulate
AB + BC = BC + CD	a) Subtraction Property of Inequality b) Substitution c) AC = BD d) Segment Addition Postulate
BC = BC	Reflexive Property of Equality
AB + CD	a) Subtraction Property of Inequality b) Substitution c) AC = BD d) Segment Addition Postulate

(Proofs Level 1, 2016).

7b) Retention answer (recognition)

Given: AC = BD
Prove: AB = CD

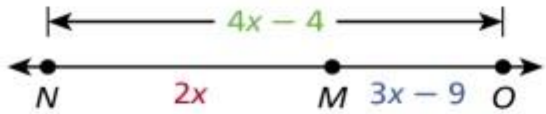


Statement	Reason
c) AC = BD	Given
AB + BC = AC	d) Segment Addition Postulate
BC + CD = BD	d) Segment Addition Postulate
AB + BC = BC + CD	b) Substitution
BC = BC	Reflexive Property of Equality
AB + CD	a) Subtraction Property of Equality

(Proofs Level 1, 2016)

8a) Transfer of Learning question (far)

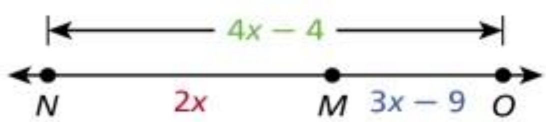
Justify how to solve the algebra equation: $4x - 4 = 2x + (3x - 9)$



Statement	Reason
$NO = NM + MO$	a) Subtraction Property of Inequality
	b) Substitution
	c) Addition Property of Equality
	d) Segment Addition Postulate
$4x - 4 = 2x + (3x - 9)$	a) Subtraction Property of Inequality
	b) Substitution
	c) Addition Property of Equality
	d) Segment Addition Postulate
$4x - 4 = 5x - 9$	Simplify
$-4 = x - 9$	a) Subtraction Property of Equality
	b) Substitution
	c) Addition Property of Equality
	d) Segment Addition Postulate
$5 = x$	a) Subtraction Property of Inequality
	b) Substitution
	c) Addition Property of Equality
	d) Segment Addition Postulate

8b) Transfer of Learning answers (far)

Justify how to solve the algebra equation: $4x - 4 = 2x + (3x - 9)$



Statement	Reason
$NO = NM + MO$	d) Segment Addition Postulate
$4x - 4 = 2x + (3x - 9)$	b) Substitution
$4x - 4 = 5x - 9$	Simplify
$-4 = x - 9$	a) Subtraction Property of Equality
$5 = x$	c) Addition Property of Equality

(Heeren, 2014).

APPENDIX B — IRB, Consent, Debrief

IRB Approval

MICHIGAN STATE
UNIVERSITY

Initial IRB Application Determination *Exempt*

January 22, 2016

To: John Smith
509C Erickson Hall

Re: **IRB# x16-030e** Category: Exempt 1
Approval Date: January 22, 2016

Title: Using Formative Feedback on Student Affect to Enhance Learning, Motivation, and Engagement

The Institutional Review Board has completed their review of your project. I am pleased to advise you that **your project has been deemed as exempt** in accordance with federal regulations.

Please submit the approval letter from Cedar Valley College upon receipt.

The IRB has found that your research project meets the criteria for exempt status and the criteria for the protection of human subjects in exempt research. **Under our exempt policy the Principal Investigator assumes the responsibilities for the protection of human subjects** in this project as outlined in the assurance letter and exempt educational material. The IRB office has received your signed assurance for exempt research. A copy of this signed agreement is appended for your information and records.

Renewals: Exempt protocols do not need to be renewed. If the project is completed, please submit an ***Application for Permanent Closure***.

Revisions: Exempt protocols do not require revisions. However, if changes are made to a protocol that may no longer meet the exempt criteria, a new initial application will be required.

Problems: If issues should arise during the conduct of the research, such as unanticipated problems, adverse events, or any problem that may increase the risk to the human subjects and change the category of review, notify the IRB office promptly. Any complaints from participants regarding the risk and benefits of the project must be reported to the IRB.

Follow-up: If your exempt project is not completed and closed after three years, the IRB office will contact you regarding the status of the project and to verify that no changes have occurred that may affect exempt status.

Please use the IRB number listed above on any forms submitted which relate to this project, or on any correspondence with the IRB office.

Good luck in your research. If we can be of further assistance, please contact us at 517-355-2180 or via email at IRB@msu.edu. Thank you for your cooperation.

Sincerely,

H. McGee

Harry McGee, MPH
SIRB Chair

c: Tim Xeriland

Office of Regulatory Affairs
Human Research Protection Programs

Biomedical & Health Institutional Review Board (BIRB)

Community Research Institutional Review Board (CRIRB)

Social Science Behavioral/Education Institutional Review Board (SIRB)

Olds Hall
408 West Circle Drive, #207
East Lansing, MI 48824
(517) 355-2180
Fax: (517) 432-4503
Email: irb@msu.edu
www.hrpp.msu.edu

Student Consent Letter

M I C H I G A N   S T A T E   U N I V E R S I T Y

COLLEGE OF EDUCATION · DEPARTMENT OF COUNSELING, EAST LANSING · MICHIGAN · 48824-1034
EDUCATIONAL PSYCHOLOGY AND SPECIAL EDUCATION

[Date]

Dear Student,

I am writing to seek your consent to participate as a research participant in *Toward Intelligent Tutor Systems that Address Student Disengagement: Can Adding Formative Reappraisal Enhance Engagement and Learning?* research project that I will direct. I am a Ph.D. candidate at Michigan State University where I work to improve mathematics teaching and learning. Also, I am a faculty member at Cedar Valley College where I can apply first-hand principles realized in research. The purpose of the project is to understand how the effects of using formative feedback on student affect impacts student engagement and learning using an intelligent tutor system (ITS). An ITS is a computer program that uses artificial intelligence to act as a tutor. Because of the large growth of these systems understanding methods that can improve student engagement while using them promises to be beneficial to school systems and those that use them. Obviously, we need the help of many students to accomplish this goal.

You must be at least 18 years old to participate in this research. During the research study, you would complete three learning units on geometric concepts. Then to conclude you would take the last assessment over the previous material covered. The study will be conducted on the Cedar Valley College campus and will take approximately three and half hours to complete but will be done over the course of three days. Individual instructors may offer extra credit for participation in this study, but those arrangements need to be made between you and your instructor. At the end of the investigation, participants that completed the study will be entered in a multiple prize raffle and the following prizes will be awarded: Samsung Galaxy Note, two HP Stream 11 laptops, and four DCCCD tuition-free 2018 spring classes with Tim Xeriland.

If you are interested in participating as a research participant, we would very much like to have you do so. If you wish to participate, please fill in the information below and click submit. If you wish not to participate, no action is required. Your participation is completely voluntary; you can choose to participate or not, without any penalty. You may refuse to answer any question, and you may also decide, at any point after you have given consent, to withdraw from research participation, without penalty. Whether you choose to participate in the study or not will not affect your grades in your classes or your standing in the college.

Should you decide to participate, we will protect your identity and your privacy during and after your participation to the maximum extent allowable by law. If your participation is to be discussed, we will assign a number or a pseudonym (a made-up name) for your work, and we would use that number (or name) in all internal and external discussions of your responses. The only people who would have access to these materials would be the members of my research team, and the Human Research Protection Program at MSU. Those that wish to receive extra credit for involvement in the study can obtain a certificate of completion upon request to show their teacher.

It is a requirement in research studies to describe potential risks and benefits to all potential participants in advance. Regarding risk, there is some possibility that you might struggle with some of the questions or be dissatisfied with your responses. Should this happen, you may terminate the project at any point. Regarding benefit, we hope that working with an ITS will be a useful learning experience for all participating students. Also, we plan to share our results with all students who choose to become research participants.

If this letter includes a sufficient description of participation, please indicate on the attached consent form if you choose to participate. If you have concerns or questions about this study, such as scientific issues, how to do any part of it, or to report an injury, please contact me, either right after the session or by phone or e-mail (below). I will respond as quickly as possible.
If you have questions or concerns about your role and rights as a research participant, would like to obtain information or offer input, or would like to register a complaint about this study, you may contact, anonymously if you wish, the Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 4000 Collins Road, Ste. 136, Lansing, MI 48910.

Sincerely,

Tim Xeriland

Tim Xeriland
3030 North Dallas Ave.
Lancaster, TX 75134-3705

(Michigan State University, 2017).

Debriefing Form for Participation in a Research Study

Cedar Valley College / Michigan State University

Thank you for your participation in our study! Your participation is greatly appreciated.

The Purpose of the Study:
We previously informed you that the purpose of the study was to determine how thinking differently about a task could improve performance. The goal of our research is to develop a better understanding of student engagement and learning while using an intelligent tutor system.

Confidentiality:
You may decide that you do not want your data used in this research. If you would like your data removed from the study and permanently deleted, please email: txeriland@dcccd.edu

Final Report:
If you would like to receive a copy of the final report of this study (or a summary of the findings) when the researcher completes it, please feel free to contact us: txeriland@dcccd.edu.

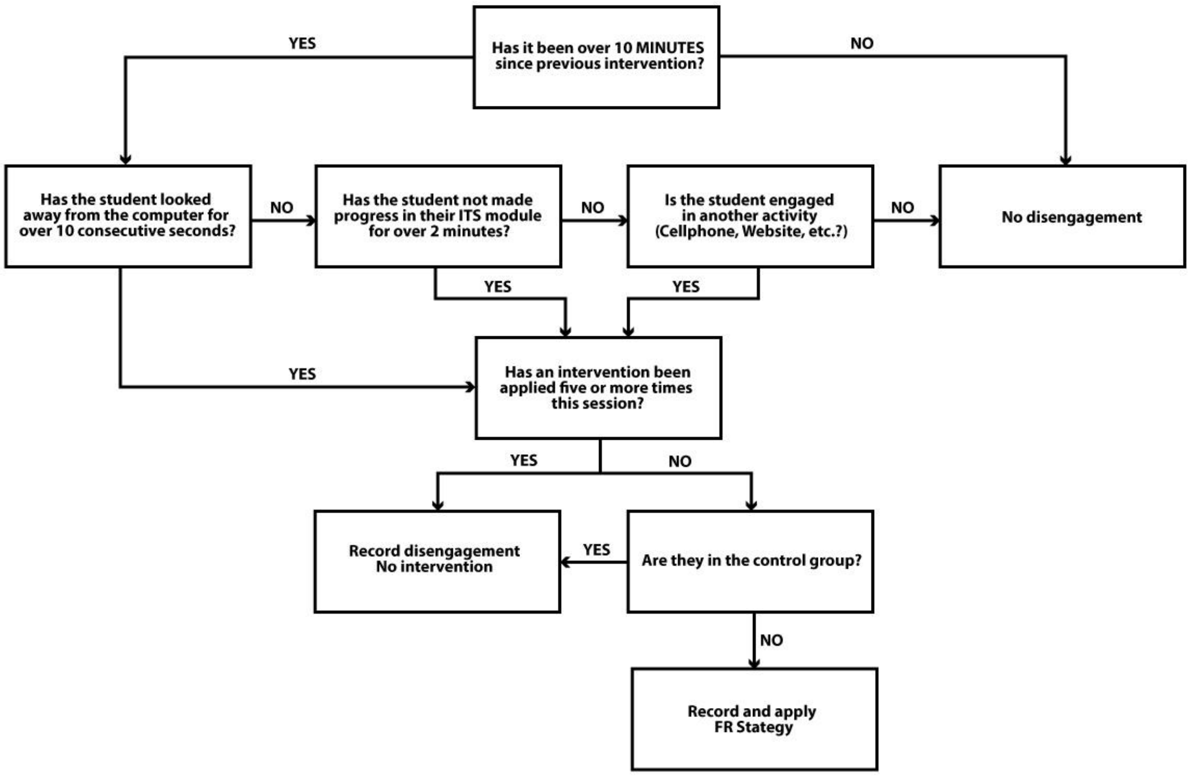
Useful Contact Information:
If you have any questions or concerns regarding this study, its purpose or procedures, or if you have a research-related problem, please feel free to contact the researcher(s), Tim Xeriland (972) 860-8239.
If you have any questions concerning your rights as a research subject, you may contact the Michigan State University's Human Research Protection Program at 517-355-2180, Fax 517-432-4503, or e-mail irb@msu.edu or regular mail at 4000 Collins Road, Ste. 136, Lansing, MI 48910.

If you feel upset after having completed the study or find that some questions or aspects of the study triggered distress, talking with a qualified clinician may help. If you feel you would like assistance, please contact Cedar Valley College's Counseling Services. Located in L- Building (L108) or call 972-860-8119.

*****Please keep a copy of this form for your future reference. Once again, thank you for your participation in this study!*****

(UMassAmherst, 2017)

APPENDIX C — Flow Chart for Determining Disengagement



REFERENCES

REFERENCES

Acee, T. W., & Weinstein, C. E. (2010). Effects of a value-reappraisal intervention on statistics students' motivation and performance. *The Journal of Experimental Education, 78*(4), 487-512. doi:10.1080/00220970903352753

Afzal, S., & Robinson, P. (2009). Natural affect data—collection & annotation in learning context. Paper presented at the Proceedings of 2009 International Conference on Affective Computing & Intelligence Interaction, Amsterdam.

Al-Radaei, S. A. M., & Mishra, R. B. (2011). A heuristic method for learning path sequencing for intelligent tutoring system (ITS). *E-learning. International Journal of Intelligent Information Technologies (IJIIT), 7*(4), 65-80. doi:10.4018/jiit.2011100104

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, *4*, 167-207.

Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education: AMEE guide no. 64. *Medical Teacher*, *34*, 148-160.

Aspinwall, L. G. (1998). Rethinking the role of positive affect in self-regulation. *Motivation and Emotion*, *22*(1), 1-32.

Auger, W. F., & Rich, S. J. (2007). *Curriculum theory and methods: Perspectives on learning and teaching*. Mississauga, Ontario: John Wiley & Sons Canada.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate learning with hypermedia? *Journal of Educational Psychology*, *96*, 523-535.

Azevedo, R., Cromley, J. G., & Seibert, D. (2004a). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, *29*, 344-370.

Azevedo, R., Guthrie, J. T., & Seibert, D. (2004b). The role of self-regulated learning in fostering students' conceptual understanding of complex system with hypermedia. *Journal of Educational Computing Research*, *30*, 87-111. doi:10.2190/DVWX-GM1T-6THQ-5WC7

Baker, R. S. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. *Proceedings of ACM CHI 2007: Computer-Human Interaction*, 1059-1068.

Baker, R. S. J. d., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human - Computer Studies, 68*(4), 223-241. doi:10.1016/j.ijhcs.2009.12.003

Baker, R. S. J. d., Moore, G., Wagner, A., Kalka, J., Karabinos, M., Ashe, C., & Yaron, D. (2011). The Dynamics between student affect and behavior occurring outside of educational software. In D'Mello, S., Graesser, A.C., Schuller, B., Martin, J.-C. (Eds.) *ACII 2011. LNCS,* (pp. 14-24). Springer-Verlag, Berlin Heidelberg.

Beal, C. R., Arroyo, I. M., Cohen, P. R., & Woolf, B. P. (2010). Evaluation of AnimalWatch: An intelligent tutoring system for arithmetic and fractions. *Journal of Interactive Online Learning, 9*, 64-77.

Beal, C. R., Walles, R., Arroyo, I., & Woolf, B. P. (2007). On-line tutoring for math achievement testing: A controlled evaluation. *Journal of Interactive Online Learning, 6,* 43-55.

Belland, B. R., Walker, A. E., & Kim, N. J. (2017). A bayesian network meta-analysis to synthesize the influence of contexts of scaffolding use on cognitive outcomes in STEM education. *Review of Educational Research, 87*(6), 1042-1081. doi:10.3102/0034654317723009

Belland, B. R., Walker, A. E., Kim, N. J., & Lefler, M. (2017). Synthesizing results from empirical research on computer-based scaffolding in STEM education: A meta-analysis. *Review of Educational Research, 87*(2), 309-344. doi:10.3102/0034654316670999

Bieg, M., Goetz, T., & Hubbard, K. (2013). Can I master it and does it matter? an intraindividual analysis on control–value antecedents of trait and state academic emotions. *Learning and Individual Differences, 28,* 102-108. doi:10.1016/j.lindif.2013.09.006

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7-74.

Blayney, P., Kalyuga, S., & Sweller, J. (2010). Interactions between the isolated–interactive elements effect and levels of learner expertise: Experimental evidence from an accountancy class. *Instructional Science, 38*(3), 277-287. doi:10.1007/s11251-009-9105-x.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*, 3-16.

Bosch, N., D'mello, S., Ocumpaugh, J., Baker, R., & Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *6*(2), 1-26. doi:10.1145/2946837

Brenner, D. G., Matlen, B. J., Timms, M. J., Gochyyev, P., Grillo-Hill, A., Luttgen, K., & Varfolomeeva, M. (2017). Modeling student learning behavior patterns in an online science inquiry environment. *Technology, Knowledge and Learning*, *22*(3), 405-425. doi:10.1007/s10758-017-9325-0

Brophy, J. E. (1981). Teacher praise: A functional analysis. *Review of Educational Research*, *51*(1), 5-32.

Brown, A. L., Bransford, J. D., Ferrara, R., & Campione, J. (1983). Learning, remembering, and understanding. In Mussen PH, Series editor. Flavell JH, Markman EM (Eds.), *Handbook of child psychology: Cognitive development*, *4th edn*. (pp. 77-166). New York, NY: Wiley.

Broza, O., & Kolikant, Y. (2015). Contingent teaching to low-achieving students in mathematics: challenges and potential for scaffolding meaningful learning, *ZDM Mathematics Education*, *47*, 1093-1105.

Burger, W. F., & Shaughnessy, J. M. (1986). Characterizing the van Hiele levels of development in geometry. *Journal for Research in Mathematics Education, 17*, 31-48.

Burleson, W., & Picard, R. (2007). Evidence for gender specific approaches to the development of emotionally intelligent learning companions. *IEEE Intelligent Systems*, *22* (4), 62-69.

Burns, H. L., & Capps, C. G. (1988). Foundations of intelligent tutoring systems: An introduction. In M. C. Polson & J.J. Richardson (Eds.), *Intelligent tutor systems* (pp. 1-19). Hillsday, NJ: Lawrence Erlbaum Associates.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1*(1), 18-37. doi:10.1109/T-AFFC.2010.1

Campuzano, L., & National Center for Education Evaluation and Regional Assistance (U.S.). (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts*. Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Dept. of Education.

Carroll, J. (1963). A model of school learning. *Teachers College Record*, *64*, 723-733.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*, 233-246.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, Academic Press, New York, 2nd edn.

Cole, J. S., Bergin, D. A., & Whitaker, T. A. (2008). Predicting student motivation for low stakes tests with effort and task value. *Contemporary Educational Psychology*, *33*, 609-624.

Cole, P. M., Michel, M. K., & Teti, L. O. (1994). The development of emotion regulation and dysregulation: A clinical perspective. *Society for Research in Child Development, 59*, 73-100.

Cooper, G., Tindall-Ford, S., Chandler, P., & Sweller, J. (2001). Learning by imagination. *Journal of Experimental Psychology: Applied, 7,* 68-82.

Corbett, A. T., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In J. Jacko, A. Sears, M. Beaudouin-Lafon, & R. Jacob (Eds.), *Proceedings of ACM CHI'2001 Conference on Human Factors in Computing Systems* (pp. 245-252). New York: The Association for Computing Machinery.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. doi:10.1016/S0022-5371(72)80001-X

Crowe, D., LaPierre, M., & Kebritchi, M. (2017). *Knowledge based artificial augmentation intelligence technology: Next step in academic instructional tools for distance learning*. New York: Springer US. doi:10.1007/s11528-017-0210-4

Crowley, M. L. (1990). Criterion-referenced reliability indices associated with the van Hiele Geometry Test. *Journal for Research in Mathematics Education, 21* (3), 238-241.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. NY: Harper-Row.

Cutuli, D. (2014). Cognitive reappraisal and expressive suppression strategies role in the emotion regulation: An overview on their modulatory effects and neural correlates. *Frontiers in Systems Neuroscience, 8*, 1-6. doi:10.3389/fnsys.2014.00175

Daniels, L. M., & Gierl, M. J. (2017). The impact of immediate test score reporting on university students' achievement emotions in the context of computer-based multiple-choice exams. *Learning and Instruction, 52*, 27-35.

De la Fuente, J., & Cardelle-Elawar, M. (2009). Research on action–emotion style and study habits: Effects of individual differences on learning and academic performance of undergraduate students. *Learning and Individual Differences, 19*(4), 567-576.

Dettmers, S., Trautwein, U., Lüdtke, O., Goetz, T., Frenzel, A. C., & Pekrun, R. (2011). Students' emotions during homework in mathematics: Testing a theoretical model of antecedents and achievement outcomes. *Contemporary Educational Psychology, 36*(1), 25-35.

Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2003). The role of feedback during academic testing: The delay retention test revisited. *The Psychological Record, 53*, 533-548.

Dillon, D. G., Ritchey, M., Johnson, B. D., & LaBar, K. S. (2007). Dissociable effects of conscious emotion regulation strategies on explicit and implicit memory. *Emotion*, *7*, 354-365.

D'Mello, S. K., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human Computer Studies*, *70*, 377-398.

D'Mello, S. K., Dieterle, E., & Duckworth, A. (2017). Advanced, Analytic, Automated (AAA) Measurement of Engagement during Learning, *Educational Psychologist, 52*(2), 104-123.

Durick, A. M., & Harackiewicz, J. M. (2007). Different strokes for different folks: How personal interest moderates the effects of situational factors on task interest. *Journal of Educational Psychology*, *99*, 597-610.

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology*, *98*, 382-393.

Dweck, C. S., Mangels, J. A., & Good, C. (2004). Motivation effects on attention, cognition, and performance. In D. Y. Dai, & R. J. Sternberg (Ed.) *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development*. Mahwah, N.J: Lawrence Erlbaum Associates. doi:10.4324/9781410610515

Eccles, J., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75-146). San Francisco, CA: W. H. Freeman.

Ehring, T., Tuschen-Caffier, B., Schnülle, J., Fischer, S., & Gross, J. J. (2010). Emotion regulation and vulnerability to depression: Spontaneous versus instructed use of emotion suppression and reappraisal. *Emotion*, *10*(4), 563-572. doi:10.1037/a0019010

Einöther, S. J. L., Baas, M., Rowson, M., & Giesbrecht, T. (2015). Investigating the effects of tea, water and a positive affect induction on mood and creativity. *Food Quality and Preference, 39*, 56–61.

El Saadawi, G. M., Azevedo, R., Castine, M., Payne, V., Medvedeva, O., Tseytlin, E., . . . Crowley, R. S. (2010). Factors affecting feeling-of-knowing in a medical intelligent tutoring system: The role of immediate feedback as a metacognitive scaffold. *Advances in Health Sciences Education*, *15*(1), 9-30. doi:10.1007/s10459-009-9162-6

Evans, M., & Michael, J. (2006). *One-on-one tutoring by humans and machines*. Mahwah, NJ: Erlbaum.

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A. A., . . . Welty, C. (2010). Building watson: An overview of the DeepQA project. *AI Magazine, 31*(3), 59.

Feyzi-Behnagh, R., Azevedo, R., Legowski, E., Reitmeyer, K., Tseytlin, E., & Crowley, R. S. (2014). Metacognitive scaffolds improve self-judgments of accuracy in a medical intelligent tutoring system. *Instructional Science, 42*(2), doi:159-181. 10.1007/s11251-013-9275-4

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-236). Hillsdale, NJ: Erlbaum.

Fong, C. J., Williams, K. M., Williamson, Z. H., Lin, S., Kim, Y. W., & Schallert, D. L. (2018). “Inside out”: Appraisals for achievement emotions from constructive, positive, and negative feedback on writing. *Motivation and Emotion, 42*(2), 236-257. doi:10.1007/s11031-017-9658-y

Foshee, C. M., Elliott, S. N., & Atkinson, R. K. (2016). Technology-enhanced learning in college mathematics remediation. *British Journal of Educational Technology, 47*(5), 893-905. doi: 10.1111/bjet.12285

Fuys, D., Geddes, D., & Tischler, R. (1988). The van hiele model of thinking in geometry among adolescents. *Journal for Research in Mathematics Education. Monograph, 3,* i-196.

Gan, S., Yang, J., Chen, X., Zhang, X., & Yang, Y. (2017). High working memory load impairs the effect of cognitive reappraisal on emotional response: Evidence from an event-related potential study. *Neuroscience Letters, 639,* 126-131. 10.1016/j.neulet.2016.12.069

Gaspard, H., Dicke, A., Flunger, B., Brisson, B. M., Häfner, I., Nagengast, B., & Trautwein, U. (2015). Fostering adolescents’ value beliefs for mathematics with a relevance intervention in the classroom. *Developmental Psychology, 51*(9), 1226-1240.

Ge, X., & Ifenthaler, D. (2017). Designing Engaging Educational Games and Assessing Engagement in Game-Based Learning. In R. Zheng, & M. Gardner (Eds.), *Handbook of Research on Serious Games for Educational Applications* (pp. 253-270). Hershey, PA: IGI Global. doi:10.4018/978-1-5225-0513-6.ch012

Germain, C. L., & Kangas, M. (2015). Trait anger symptoms and emotion regulation: The effectiveness of reappraisal, acceptance and suppression strategies in regulating anger. *Behaviour Change*, *32*(1), 35-45.

Godes, O., Hulleman, C. S., & Harackiewicz, J. M. (2007, April). Boosting students’ interest in math with utility value: Two experimental tests. Paper presented at the annual conference of the American Educational Research Association, Chicago, IL.

Goldin, P. R., McRae, K., Ramel, W., & Gross, J. J. (2008). The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biological Psychiatry, 63*, 577-586.

Graesser, A. C., VanLehn, K., Rose, C. P., Jordan, P. W., & Harter, D. (2001). Intelligent tutoring systems with conversational dialogue. *AI Magazine, 22*(4), 39.

Greene, B., & Land, S. (2000). A qualitative analysis of scaffolding use in a resource-based learning environment involving the World Wide Web. *Journal of Educational Computing Research, 23*(2), 151-179.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271-299.

Gross, J. J. (2008). Emotion regulation: Personality processes and individual differences. In O.P. John, R. W. Robbins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 701-722). New York, NY: Guilford Pres.

Gross, J. J. (2014). Emotion regulation: Conceptual and empirical foundations. *Handbook of Emotion Regulation* (2nd ed. pp. 3-20). New York: Guilford Press.

Gross, J. J., Feldman Barrett, L., & Richards, J. M. (1998). *Emotion regulation in everyday life*. Manuscript in preparation.

Gross, J. J., & John, O. P. (1998). Mapping the domain of emotional expressivity: Multi-method evidence for a hierarchical model. *Journal of Personality and Social Psychology, 74*, 170-191.

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*(2), 348-362. doi:10.1037/0022-3514.85.2.348

Gross, J. J., & Thompson, R. (2007). Emotion regulation: Conceptual foundations. In J. J. Gross (Ed.), *Handbook of Emotion Regulation* (pp. 3-24). New York: Guilford Press.

Gumora, G., & Arsenio, W. F. (2002). Emotionality, emotion regulation, and school performance in middle school children. *Journal of School Psychology*, *40*(5), 395-413.

Guzmán, E., Conejo, R., & Pérez-de-la-Cruz, J. (2007). Adaptive testing for hierarchical student models. *User Modeling and User-Adapted Interaction, 17*(1), doi:119-157. 10.1007/s11257-006-9018-1

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. 10.3102/003465430298487

Hayes, J. P., Morey, R. A., Petty, C. M., Seth, S., Smoski, M. J., McCarthy, G., & Labar, K. S. (2010). Staying cool when things get hot: Emotion regulation modulates neural mechanisms of memory encoding. *Frontiers in Human Neuroscience, 4*, 230. doi:10.3389/fnhum.2010.00230

Heeren, J. (2014). Sophia: Algebraic Proof. Unpublished instrument. Retrieved from https://www.sophia.org/tutorials/algebraic-proof

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist, 41*, 111-127.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*(3), 319-340. doi:10.1037/0033-295X.94.3.319

Howes, M. B., & O'Shea, G. (2014). *Human memory: A constructivist view*. San Diego: Academic Press.

Hulleman, C. S., Durik, A. M., Schweigert, S. A., & Harackiewicz, J. M. (2010a). Task Values, Achievement Goals, and Interest: An Integrative Analysis. *Journal of Educational Psychology, 100*, 398-416.

Hulleman, C. S., Godes, O., Hendricks, B. L., & Harackiewicz, J. M. (2010b). Enhancing interest and performance with a utility value intervention. *Journal of Educational Psychology, 102*, 880-896.

IBM Pearson. (2016). IBM Watson Education and Pearson to drive cognitive learning experiences for college students [press release]. Somers: IBM Corporation Retrieved from http://www-03.ibm.com/press/us/en/pressrelease/50842.wss

Introducing Geometry Proofs - A New Approach. (n.d.). Retrieved July 20, 2017, from http://www.mathgiraffe.com/blog/introducing-two-column-geometry-proofs-a-different-approach

James, W. (1890). *The principles of psychology*. New York: H. Holt and company.

Jeremić, Z., Jovanović, J., & Gašević, D. (2012). Student modeling and assessment in intelligent tutoring of software patterns. *Expert Systems with Applications, 39*(1), 210-222. doi:10.1016/j.eswa.2011.07.010

John, O. P., & Gross, J. J. (2004). Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. *Journal of Personality, 72*(6), doi:1301-1334. 10.1111/j.1467-6494.2004.00298.x

Johnson-Laird, P. N., & Mancini, F. (2006). A hyper-emotion theory of psychological illnesses. *Psychological Review, 113* (4), 822-841.

Jones, D. M., & Macken, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*, 369-381.

Kalyuga, S. (2008). When Less Is More in Cognitive Diagnosis: A Rapid Online Method for Diagnosing Learner Task-Specific Expertise. *Journal of Educational Psychology, 100*(3), 603-612.

Kalyuga, S., & Renkl, A. (2010). Expertise reversal effect and its instructional implications: Introduction to the special issues. *Instructional Science, 38*, 209-215.

Kalyuga, S., Rikers, R., Pass, F. (2012). Educational implications of expertise reversal effect in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review, 24*, 313-337.

Karam, R., Pane, J. F., Griffin, B. A., Robyn, A., Phillips, A., & Daugherty, L. (2017). Examining the implementation of technology-based blended algebra I curriculum at scale. *Educational Technology Research and Development*, *65*(2), 399-425. doi:10.1007/s11423-016-9498-6

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19* (3), 239–264.

Koedinger, K. R., Brunskill, E., Baker, R. S., McLaughlin, E. A., & Stamper, J. (2013). New potentials for data-driven intelligent tutoring system development and optimization. *AI Magazine, 34* (3), 1-17.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research, 86*(1), 42-78. doi:10.3102/0034654315581420

Kulik, C.-L. C., & Kulik, J., A. (1984, August). Effects of ability grouping on elementary school pupils: A meta-analysis. Paper presented at the annual meeting of the American Psychological Association, Toronto, Canada. (ERIC Document Reproduction Service No. ED255329).

Kulik, C.-L. C., & Kulik, J. A. (1985, August). *Effects of ability grouping on achievement and self-esteem*. Paper presented at the annual convention of the American Psychological Association, Los Angeles, CA.

Kulik, C-L. C., & Kulik, J. A. (1991a). Effectiveness of computer-based instruction: An updated analysis. *Computers in Human Behavior, 7*, 75-91.

Kulik, J. A., & Kulik, C-L. C. (1991b). Ability grouping and gifted students. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (pp. 178-196). Boston, MA: Allyn & Bacon.

Kulik, J. A., & Kulik, C-L. C. (1992). Meta-analytic findings on grouping programs. Gifted *Child Quarterly, 36*, 73-77.

Larcom, M. J., & Isaacowitz, D. M. (2009). Rapid emotion regulation after mood induction: Age and individual differences. Journal of Gerontology: *Psychological Sciences, 64,* 733-741.

Lazarus, R. S., & Alfert, E. (1964). Short-circuiting of threat by experimentally altering cognitive appraisal. *The Journal of Abnormal and Social Psychology, 69*(2), 195-205. doi:10.1037/h0044635

Leahy, W., & Sweller, J. (2004). Cognitive load and the imagination effect. *Applied Cognitive Psychology, 18*(7), 857–875.

Leff, L. S. (1997). *Geometry the easy way*. Hauppauge, NY: Barron's Educational Series.

Leahy, W., & Sweller, J. (2005). Interactions Among the Imagination, Expertise Reversal, and Element Interactivity Effects. *Journal of Experimental Psychology: Applied*. *11*(4), 266-276.

Lepper, M. R., Woolverton, M., Mumme, D. L., & Gurtner, J. (1993). Motivational Techniques of Expert Human Tutors: Lessons for the Design of Computer-Based Tutors. In Lajoie, S.P. & Derry, S.J. (Eds.) *Computers as cognitive tools* (pp. 75-105). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Lin, C. C., Guot, K. H., & Lin, Y. C. (2016). A simple and effective remedial learning system with a fuzzy expert system. *Journal of Computer Assisted Learning, 32*(6), 647-662.

Linnenbrink, E. A. (2006). Emotion research in education: Theoretical and methodological perspectives on integration of affect, motivation, and cognition. *Educational Psychology Review*, *18*(4), 307-314.

Linnenbrink, E. A. (2007). The role of affect in student learning-chapter 7: A multi-dimensional approach to considering the interaction of affect, motivation, and engagement. In *Emotion in Education* (pp. 107-124). Ontario, Canada: Elsevier Inc. doi:10.1016/B978-012372545-5/50008-3

Linnenbrink, E. A., Pintrich, P. R. (2002) The role of motivational beliefs in conceptual change. In: M. Limón, L. Mason (Eds.), *Reconsidering conceptual change: Issues in theory and practice* (pp. 115-135). Netherlands: Springer, Dordrecht.

Linnenbrink, E. A., & Pintrich, P. R. (2004). Role of affect in cognitive processing in academic contexts. In Dai DY, Sternberg RJ, (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development*, (pp. 57-87). Mahwah, N.J: Lawrence Erlbaum Associates. doi:10.4324/9781410610515

Loveless, T. (2015). *How well are Americans students learning?* Washington DC: Brookings Institution.

Lucast, E. K. (2003). Proof as Method: A New Case for Proof in Mathematics Curricula. Unpublished masters thesis. Pittsburgh, PA, USA: Carnegie Mellon University.

Manjoo, F. (2018, February 27). The sublime and scary future of cameras with AI brains. *New York Times.* Retrieved https://nyti.ms/2t3i5oN

McRae, K., Ciesielski, B., & Gross, J. J. (2012). Unpacking cognitive reappraisal: Goals, tactics, and outcomes. *Emotion (Washington, D.C.), 12*(2), 250-255. doi:10.1037/a0026351

Menager-Beeley, R. (2001). Student success in web based distance learning: Measuring motivation to identify at risk students and improve retention in online classes. *WebNet 2001: World Conference on the WWW and Internet Proceedings*, Orlando, FL.

Michigan State University. (2017). Guidance document – Office of Research Ethics and Standards Human Research Protection Program. Retrieved from https://hrpp.msu.edu/sites/default/files/12-4%20Consent%20Form%20Guidelines_12-12-2015.pdf

Midgley, C., Maehr, M. L., Hruda, L., Anderman, E. M., Anderman, L., Freeman, K. E., … Urdan, T., (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor, MI: University of Michigan.

Najar, A. S., Mitrovic, A., & McLaren, B. M. (2016). Learning with intelligent tutors and worked examples: Selecting learning activities adaptively leads to better learning outcomes than a fixed curriculum. *User Modeling and User-Adapted Interaction, 26*, 459-491. doi:10.1007/s11257-016-9181-y

Nahadi, Firman, H., & Farina, J. (2015). effect of feedback in formative assessment in the student learning activities on chemical course to the formation of habits of mind. *Jurnal Pendidikan IPA Indonesia*, *4*(1), 36-42. doi:10.15294/jpii.v4i1.3499

Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Ed.), *Instructional design for multimedia learning* (pp. 181–195). Munster, New York: Waxmann.

Nevid, J. S. (2007). Kant, cognitive psychotherapy, and the hardening of the categories. *Psychology and Psychotherapy: Theory, Research and Practice. 80* (Pt 4): 605–615. doi:10.1348/147608307X204189. Or 10.1348/147608307X204189

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in Higher Education, 31(2), 199-218. doi:10.1080/03075070600572090

Ortner, C. N. M., Ste Marie, M., & Corno, D. (2016). Cognitive costs of reappraisal depend on both emotional stimulus intensity and individual differences in habitual reappraisal. *Plos One, 11*(12), 1-16. doi:e0167253. 10.1371/journal.pone.0167253

Padesky, C. A. (1994). Schema change processes in cognitive therapy. *Clinical Psychology and Psychotherapy 1*(5), 267-278.

Parlett, D. K. (2012). A comparison of associate and bachelor degree nursing students' motivation (Unpublished doctoral dissertation). Walden University, Minneapolis, MN.

Pekrun, R. (2000). A social cognitive, control-value theory of achievement emotions. In J. Heckhausen (Ed.), *Motivational psychology of human development* (pp. 143-163). Oxford: Elsevier.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315-341.

Pekrun, R. (2009). Emotions at school. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 575-604). New York: Routledge.

Pekrun, R. (2017). Emotion and achievement during adolescence. *Child Development Perspectives, 11*(3), 215-221. doi:10.1111/cdep.12237

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz & R Pekrun (Eds.), *Emotion in education*, (pp. 13-36). San Diego, CA: Academic Press.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*(3), 531-549.

Pekrun, R., Goetz, T., Frenzel, A., Barchfeld, P., & Perry, R. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*, 36-41. doi:10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., & Hochstadt, M. (2004). Beyond test anxiety: Development and validation of the Test Emotions Questionnaire (TEQ). *Anxiety, Stress and Coping*, *17*, 287-316.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educ Psychol 37*, 99-105.

Pekrun, R., & Linnenbrink-Garcia, L. (2014). Introduction to emotions in education. In R. Pekrun & L. Linnenbrink-Garcia (Eds.) *International handbook of emotions in education* (pp. 1-10). New York: Taylor & Francis.

Phye, G. D., & Andre, T. (1989). Delayed retention effect: Attention, perseveration, or both? *Contemporary Educational Psychology*, *14*(2), 173–185.

Piaget, J. (1976). Piaget’s theory. In B. Inhelder, H.H. Chipman & C. Zwingman (Coordinating Ed). *Piaget and his school: A reader in developmental psychology*. New York, NY: Springer.

Piaget, J., & Cook, M. T. (1952). *The origins of intelligence in children*. New York, NY: International University Press.

Pintrich, P. R. (1995). *Understanding self-regulated learning*. San Francisco: Jossey-Bass.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: National Center for Research to Improve Post-Secondary Teaching.

PISA. (2012). Retrieved December 3, 2015, from http://www.oecd.org/pisa/keyfindings/PISA-2012-results-US.pdf

Popham, W. J. (2011). *Transformative assessment in action: An inside look at applying the process.* Alexandria: ASCD.

Proofs Level 1. (2016). Retrieved March 26, 2016, from http://www.pleasanton.k12.ca.us/avhsweb/kiyoi/OnlinePracticeProofs/Proofs_Level_1B.html

Putwain, D. W., & Symes, W. (2012). Achievement Goals as Mediators of the Relationship between Competence Beliefs and Test Anxiety. *British Journal of Educational Psychology*, *82*(2), 207-224.

Rebello, N. S., Cui, L., Bennett, A. G., Zollman, D. A., & Ozimek, D. J. (2007). Transfer of learning in problem solving in the context of mathematics and physics, in *Learning to Solve Complex Scientific Problems*, D.H. Jonassen, Editor. 2007, Lawrence Erlbaum Associates: Mahwah, NJ.

Reif, F., & Scott, L. A. (1999). Teaching scientific thinking skills: Students and computers coaching each other. *American Journal of Physics, 67*, 819–831.

Reisenzein, K. (2001), Appraisal processes conceptualized from a schema-theoretic perspective, In K, R, Scherer. A Schorr. & T. Johnstone, 1. (Eds.), *Appraisal processes in emotion* (pp, 187-201). Oxford, UK: Oxford University Press.

Renkl, A., & Atkinson, R. K. (2007). Interactive learning environments: Contemporary issues and trends. An introduction to the special issue. *Educational Psychology Review, 19*(3), 235-238.

Rey, G. D., & Buchwald, F. (2011). The expertise reversal effect: Cognitive load and motivational explanations. *Journal of Experimental Psychology Applied 17*(1), 33-48.

Richards, J. M., & Gross, J. J. (1999). Composure at any cost? The cognitive consequences of emotional suppression. *Personality and Social Psychology Bulletin*, *25*, 1033-1044.

Richards, J. M., & Gross, J. J. (2000). Emotion regulation and memory: The cognitive costs of keeping one's cool. *Journal of Personality and Social Psychology, 79*(3), 410-424. 10.1037//0022-3514.79.3.410

Rogers, K. B. (1991). *The relationship of grouping practices to the education of the gifted and talented learner* (RBDM9102). Storrs: University of Connecticut, The National Research Center on the Gifted and Talented.

Rosenbloom, T., Shahar, A., Perlman, A., Estreich, D., & Kirzner, E. (2007). Success on a practical driver's license test with and without the presence of another testee. *Accident Analysis & Prevention. 39*(6), 1296-1301.

Rosner, D., Haase, M., Bauer, T., Gunther, T., Kruger, J., & Frommer, J. (2015). Desiderata for the design of companion systems: Insights from a large scale Wizard of Oz experiment. *Künstliche Intelligenz*, *12*, 131-157.

Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In: R.J. Spiro et al. (eds.) *Theoretical Issues in Reading Comprehension*, Hillsdale, NJ: Lawrence Erlbaum.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*, 493-502.

Sabourin, J., Rowe, J., Mott, B., & Lester, J. (2011). When off-task in on-task: The affective role of off-task behavior in narrative-centered learning environments. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. *LNCS*, vol. 6738, pp. 534-536. Springer, Heidelberg.

SACS (2018). *Quality enhancement plan: Cedar valley college*. Atlanta, GA: Southern Association of Colleges and Schools.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*: 119–144.

Salden, R., Aleven, V., Renkl, A., & Schwonke, R. (2008). Worked examples and tutored problem solving: Redundant or synergistic forms of support? In C. Schunn (Ed.) *Proceedings of the 30th Annual Meeting of the Cognitive Science Society, CogSci* 2008 (pp. 659-664). New York, NY: Lawrence Earlbaum.

Salden, R. J., C., M., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science, 38*(3), 289-307. http://dx.doi.org.proxy2.cl.msu.edu/10.1007/s11251-009-9107-8

Schutz, P., & Davis, H. A. (2010). Emotion regulation related to a particular test. *Proceedings of the 11th International Conference on Education Research New Educational Paradigm for Learning and Instruction* (pp. 57-59). Seoul, South Korea: Seould National University.

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education: Theory, research, and applications* (3rd ed.). Upper Saddle River, N.J: Pearson/Merrill Prentice Hall.

Schwartz, F., & White, K. (2000). Making sense of it all: Giving and getting online course feedback. In K. W. White & B. H. Weight (Eds.), *The online teaching guide: A handbook of attitudes, strategies, and techniques for the virtual classroom* (pp. 57-72). Boston: Allyn and Bacon.

Schwonke, R., Renkl, A., Salden, R., & Aleven, V. (2011). Effects of different ratios of worked solution steps and problem solving opportunities on cognitive load and learning outcomes. *Computers in Human Behavior, 27*(1), 58-62. 10.1016/j.chb.2010.03.037

Schofield, J. W. (1995). *Computers and classroom culture*. Cambridge: Cambridge University Press.

Senk, S. (1989). Van Hiele levels and achievement in writing geometry proofs. *Journal for Research in Mathematics Education, 20*, 309-321.

Sheldon, K. M., Ryan, R. M., Rawsthorne, L. J., & Ilardi, B. (1997). Trait self and true self: Cross-role variation in the big-five personality traits and its relations with psychological authenticity and subjective well-being. *Journal of Personality and Social Psychology, 73*(6), 1380-1393. doi:10.1037/0022-3514.73.6.1380

Sheppes, G., & Meiran, N. (2007). Better late than never? on the dynamics of online regulation of sadness using distraction and cognitive reappraisal. *Personality and Social Psychology Bulletin, 33*(11), 1518-1532. doi:10.1177/0146167207305537

Sheppes, G., & Meiran, N. (2008). Divergent cognitive costs for online forms of reappraisal and distraction. *Emotion, 8*(6), 870-874. doi:10.1037/a0013711

Shiota, M. N., & Levenson, R. W. (2009). Effects of aging on experimentally instructed detached reappraisal, positive reappraisal, and emotional behavior suppression. *Psychology of Aging*, *24*(4), 890-900.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153-189.

Siegler, R. S., DeLoache, J. S., & Eisenberg, N. (2003). *How children develop*. New York: Worth.

Simons, J., Vansteenkiste, M., Lens, W., & Lacante, M. (2004). Placing motivation and future time perspective theory in a temporal perspective. *Educational Psychology Review*, *16*, 121–139.

Skinner, B. F. (1958). Teaching machines. *Science, 128*, 969-977. doi:10.1126/science.128.3330.969

Smith, S. G., & Sherwood, B. A. (1976). Educational uses of the PLATO computer system. *Science, 192*, 344-352.

Sparks, S. D. (2017). How 'intelligent' tutors could transform teaching; teachers' jobs aren't going away, but they could be different. *Education Week, 37*(6), 1-3.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology, 105*, 970-987. doi:10.1037/a0032447

Strain, A. C., & D'Mello, S. K. (2015). Affect regulation during learning: The enhancing effect of cognitive reappraisal. *Applied Cognitive Psychology*, *29*, 1-19.

Sun, M. (2014). Retrieved March 5, 2018, from https://www.psychologyinaction.org/psychology-in-action-1/2014/02/17/psychological-skills-changing-your-emotions-an-intro-to-cognitive-reappraisal

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*. *12*(2), 257-285.

Sweller, J. (1999). *Instructional Design in Technical Areas*, Camberwell, Victoria, Australia, Australian Council for Educational Research.

Sweller, J. (2005). Implications of cognitive load for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). Cambridge, MA: Cambridge University Press.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59-89.

Taber, K. S. (2010). Paying lip-service to research? the adoption of a constructivist perspective to inform science teaching in the English curriculum context. *Curriculum Journal, 21*(1), 25-45. doi:10.1080/09585170903558299

The Transitive and Substitution Properties. (n.d.). Retrieved July 20, 2017, from http://www.dummies.com/education/math/geometry/the-transitive-and-substitution-properties/

Trevors, G., Duffy, M., & Azevedo, R. (2014). Note-taking within MetaTutor: Interactions between an intelligent tutoring system and prior knowledge on note-taking and learning. *Educational Technology Research and Development*, *62*(5), 507-528. doi:10.1007/s11423-014-9343-8

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *The American Journal of Psychology*, *9*(4), 507-533.

UMassAmherst. (2017). Research administration & compliance: Debriefing form – deception. Retrieved from https://www.umass.edu/research/form/debriefing-form-deception.

Usiskin, Z. (1982). *Van Hiele levels and achievement in secondary school geometry*. Chicago: University of Chicago, Cognitive Development and Achievement in Secondary School Geometry Project.

Usiskin, Z., & Senk, S. (1990). Evaluating a test of van hiele levels: A response to crowley and wilson. *Journal for Research in Mathematics Education, 21*(3), 242-245. doi:10.2307/749378

Van Hiele, P. M. (1980). *Levels of thinking: How to meet them and how to avoid them.* Paper presented at the research presentation to the annual meeting of the National Council of Teachers of Mathematics, Seattle, WA.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist, 46*(4), 197-221.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science, 31*(1), 3-62.

Wiliam, D. (1999). Formative assessment in mathematics part 2: Feedback. *Equals: Mathematics and Special Educational Needs, 5*(3) 8-11.

Winne, P. H., & Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology*, *28*(3), 259-276. doi:10.1016/S0361-476X(02)00041-3

Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review, 20*(4), 429-444. doi:10.1007/s10648-008-9080-9

Woolf, B. P. (2009). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Burlington, MA: Morgan Kaufmann.

Xeriland, T. J. (2012). The effects of formative evaluation on learning mathematics online (Unpublished Practicum). Michigan State University, Lansing, MI.

Yang, Y. C., Gamble, J. H., Hung, Y., & Lin, T. (2014). An online adaptive learning environment for critical-thinking-infused English literacy instruction. British Journal of Educational Technology, 45(4), 723-747. doi:10.1111/bjet.12080

Yarnall, L., Means, B., & Wetzel, T. (2016). Lessons learned from early implementations of adaptive courseware. Retrieved from SRI International website: https://www.sri.com/sites/default/files/brochures/almap_final_report.pdf

Zajonc, R. B. (1965). Social Facilitation. *Science. American Association for the Advancement of Science*. *149*(3681), 269-274.