

Promoting Self-Regulation and Metacognition through the Use of Online Trace Data
within a Game-Based Environment

by

Erica L. Snow

A Dissertation in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2015 by the
Graduate Supervisory Committee:

Danielle McNamara, Chair
Carol Connor
Nicholas Duran
Phillip Winne

ARIZONA STATE UNIVERSITY

December 2015

ABSTRACT

Computer-based environments provide a window into the complex and multifaceted learning process. These systems often collect a vast amount of information concerning how users choose to engage and behave within the interface (i.e., click streams, language input, and choices). Researchers have begun to use this information to gain a deeper understanding of users' cognition, attitudes, and abilities. This dissertation is comprised of two published articles that describe how post-hoc and real-time analyses of trace data provides fine-grained details about how users regulate, process, and approach various learning tasks within computer-based environments. This work aims to go beyond simply understanding users' skills and abilities, and instead focuses on understanding *how* users approach various tasks and subsequently using this information in *real-time* to enhance and personalize the user's learning experience.

DEDICATION

This dissertation is dedicated to my best friend and biggest supporter, Zach Snow. I could have never achieved my dreams without your unwavering support. Thank you.

ACKNOWLEDGEMENTS

The research presented in this dissertation was supported in part by the Institute for Educational Sciences (IES R305G020018-02, R305G040046, R305A080589) and National Science Foundation (NSF REC0241144, IIS-0735682). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES or NSF. I am extremely grateful to Dr. Danielle McNamara who has thoughtfully guided my development as a scientist throughout my entire career. I am also grateful to the many members of the SoLET Lab who I have had the honor of working with these past four years.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER	
INTRODUCTION	1
Chapter 1 Synopsis	1
Chapter 2 Synopsis	2
General Discussion	3
1. DOES AGENCY MATTER?: EXPLORING THE IMPACT OF CONTROLLED BEHAVIORS WITHIN A GAME-BASED ENVIRONMENT	4
Introduction	4
Current Study	10
iSTART-2	11
Study 1	15
Study 1 Method	16
Quantitative Method	20
Study 1 Results	27
Study 1 Discussion	28
Study 2	31
Study 2 Method	32
Study 2 Results	34
Study 2 Discussion	37

CHAPTER	Page
Conclusion.....	39
References	43
2. PROMOTING METACOGNITION IN A GAME-BASED STRATEGY TUTOR.....	50
Introduction	50
iSTART-2	54
Current Study.....	56
Study 1	59
Study 1 Method	60
Study 1 Results	62
Study 1 Discussion	65
Study 2	66
Study 2 Method	66
Study 2 Results	67
Study 2 Discussion	72
Study 3	73
Study 3 Method	73
Study 3 Results	73
Study 3 Discussion	78
Conclusion.....	80
References	83
3. OVERALL DISCUSSION	86
COMPREHENSIVE REFERENCES	92

LIST OF TABLES

Table	Page
1. Posttest Attitude Measures.....	20
2. Vector Assignments	22
3. Posttest Performance Threshold Feature Perception Questions.....	67
4. Posttest Performance Threshold Feature Perceptions	72
5. Posttest Transition Message Perceptions per Condition	77

LIST OF FIGURES

Figure	Page
1. Screen Shot of iSTART-2 Selection Menu	12
2. Screen Shot of Showdown	13
3. Screen Shot of Balloon Bust	15
4. Example Random Walk	22
5. Actual Random Walks from Study 1	23
6. Actual Random Walks from Study 1	23
7. Example of Controlled Pattern (i.e., Low Entropy)	26
8. Example of Disordered Pattern (i.e., High Entropy)	26
9. Self-Explanation Scores across Studies 1 and 2	37
10. Interface of iSTART-2	55
11. Screen Shot of Map Conquest	55
12. Screen Shot of Performance Threshold Pop-up Message	57
13. Screen Shot of Coached Practice	59
14. Transition Gain Scores per Student Study 1	64
15. Transition Gain Scores per Student Study 2	69

INTRODUCTION

The current work is comprised of two chapters designed to examine two main research questions, 1) can online data be used to understand how students regulate their behaviors within the game-based environment iSTART-2, and 2) can information gleaned from online data be used in real-time to provide guidance and promote reflection during learning tasks. The work presented in this dissertation is unique in that it uses both post-hoc and real-time trace data analysis to understand and promote self-regulative processes within a game-based environment. Specifically, in Chapter 1, two studies are conducted to examine how post-hoc analyses of students' choice patterns may be indicative of the amount of agency that the student is exerting during learning. Chapter 2 builds off of this work through the development and testing of a new game-based feature that uses information gleaned from students' real-time performance data to provide guidance and promote reflection during learning tasks within iSTART-2. By investigating both post-hoc and real-time patterns of interactions and performance, we may be able to gain a deeper understanding of how students regulate their behaviors within adaptive systems and how various features can be built into these systems to promote optimal learning behaviors.

CHAPTER 1 SYNOPSIS

When students exhibit control and employ a strategic plan of action over a situation they are said to be demonstrating agency (Bandura, 2001). Chapter 1, published in *Computers and Education* (Snow, Allen, Jacovina, & McNamara, 2015), is comprised of two studies designed to investigate how agency manifests within students' choice patterns and

ultimately influences self-explanation quality within iSTART-2. In Study 1, 75 college students interacted freely within iSTART-2 for two hours. Random walk and Entropy analyses were used to quantify the amount of control demonstrated in students' choice patterns, as well as to determine the relation between variations in these patterns and self-explanation performance within iSTART-2. Overall, students who demonstrated more controlled choice patterns generated higher quality self-explanations compared to students who exhibited more disordered choice patterns. This link between performance and controlled choice patterns is hypothesized to be driven, in part, by students' experiences of agency. That is, engaging in controlled patterns should be advantageous only when doing so is a result of students' strategic planning. In Study 2, this hypothesis was tested by assigning 70 students to a choice pattern (i.e., controlled or disordered) that had been yoked to students from Study 1, thus removing students' ability to exert agency over the iSTART-2 system. Results revealed no differences in self-explanation quality between the groups assigned to controlled and disordered choice patterns. Collectively, findings from these studies support the notion that success within game-based systems is related to students' ability to exert agency over their learning paths.

CHAPTER 2 SYNOPSIS

Metacognitive skills have been shown to be critical for academic success. However, students often struggle to regulate these skills during learning tasks. Chapter 2 (Snow, Jacovina, Allen, & McNamara, under review at *Computers in Education*) is comprised of three studies that investigate how features designed to promote metacognitive awareness can be built into iSTART-2. Across the three studies, 140 college students interacted with iSTART-2 for one hour, completing lesson videos and practice activities. If students'

performance fell below a minimum threshold during game-based practice, they received a pop-up that alerted them of their poor performance and they were subsequently transitioned to a remedial activity. Results revealed that students' in-system performance improved after they were transitioned. Further, these results did not vary as a function of individual differences in prior knowledge. Thus, both high and low ability students benefited from this performance threshold feature built into iSTART-2. These results suggest that the performance threshold feature indirectly promotes metacognition, thus leading to increased performance. Overall, these studies provide insight into the potential benefits of real-time feedback designed to promote metacognitive awareness within a game-based learning environment.

GENERAL DISCUSSION

This dissertation work examines how information gleaned from online process data can be used to understand aspects of students' ability to self-regulate along with how this information can be used to guide adaptive instruction and prompt self-reflection during learning tasks. Combined findings from Chapter 1 (Snow et al., 2015) and Chapter 2 (Snow et al., under review) provide evidence regarding the importance of analyzing and quantifying variations in students' interactions within game-based environments. The two completed studies presented within this document are among of the first to use trace-data from game-based environments to investigate *how* users approach various learning tasks and then use information gleaned from that data to adapt system pedagogy. Thus, the current work may serve as a starting point for scientists interested in analyzing how interactions within game-based environments provide a deeper understanding of the various aspects of students' regulatory processes.

CHAPTER 1

“Does Agency Matter?: Exploring the Impact of Controlled Behaviors within a Game-Based Environment”

Published in the Journal of Computers and Education

Everyday people make decisions, set plans, and exert influence over their daily lives. Individuals control a multitude of situations through their choices, decisions, and strategic plans. They exert *agency* (Metcalfe, Eich, & Miele, 2013). Indeed, agency (or lack thereof) is a pervading aspect of our lives.

According to Bandura (2001), there are four main features of human agency: intentionality, forethought, self-reactiveness, and self-reflectiveness. *First*, agency refers to a *deliberate* action or set of actions, which are purposefully enacted according to a specific plan. *Second*, agency involves forethought. A person exerting high levels of agency will have set goals and plans for how to obtain these goals before carrying out any actions. *Third*, agency is self-reactive, emerging from a person’s motivation to succeed and self-regulate. *Finally*, agency is self-reflective, such that the person who is exerting agency is metacognitively aware of the goals, plans, and behavior adjustments necessary to complete a task. Combined, these four components portray agency as a dynamic behavior that involves intentionality, metacognition, and planned sets of behaviors designed to accomplish a goal. Considering agency in this light, it is not surprising that individuals who demonstrate agency over their environment generally lead more successful lives compared to those people who do not (Bandura, 2001; Ford, 1992).

Educational Environments and Student Agency

Within the realm of education, agency emerges as an important factor that influences students' engagement and subsequent learning of academic material (Bandura, 1989; Zimmerman, 2008). Indeed, it is a widely accepted belief in the classroom that affording students control promotes their motivation and subsequent learning outcomes (Flowerday & Schraw, 2000). In support of that assumption, students' emotions while learning have been moderately associated with their perceptions of subjective control (Pekrun, 2006; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010). According to the control-value theory of emotion (Pekrun, 2006; Pekrun et al., 2010), a greater sense of control is related to students' expressed positive attitudes during a learning task. Consequently, engagement is expected to increase when students are afforded autonomy and control during various academic tasks (Calvert, Strong, & Gallagher, 2005; Cordova & Lepper, 1996; Deci & Ryan, 1985; Hidi & Renninger, 2006). Cordova and Lepper (1996), for instance, found that learners who were afforded more opportunities to exert control over a learning task reported stronger motivation and interest and showed better performance on a subsequent math test. Similarly, Calvert and colleagues (2005) reported that students who were given more control over a computer-based storybook reported greater interest in the task and were more attentive to the material than students who were given more explicit directions by adults. These studies provide growing evidence that agency during learning has the potential to enhance motivation, interest, and attitudes, all of which are associated with positive learning outcomes.

Adaptive learning environments have attempted to leverage these positive effects of agency by incorporating various elements such as customization, games, and “choose

your own adventures.” These game-based elements are designed to promote students’ feelings of agency and by consequence enhance motivation, performance, and learning outcomes (Cordova & Lepper, 1996; Jackson & McNamara, 2013; Snow, Jackson, Varner, & McNamara, 2013). For example, providing learners with control over when and how long they engage with different lessons (i.e., their learning trajectory) in a system can improve learning outcomes (Tabbers & de Koeijer, 2010). Even giving control over educationally superficial features of a system (e.g., choosing the images that will be depicted by the system) allows learners’ experiences to match their personal preferences. This in turn can decrease the effort required to engage in a task and subsequently increase involvement and learning (Corbalan, Kester, & van Merriënboer, 2009; 2011). In this way, students can feel as though they are exerting control over their environment with minimal changes to the learning task itself.

Game-based systems are particularly germane to the issue of a student’s sense of agency during learning. By leveraging the mechanics and features found in popular, non-educational games, learning environments infused with games naturally afford students the ability to exert influence on the learning environment (McNamara, Jackson, & Graesser, 2010). Indeed, a number of game features have been adapted from popular games to educational games with the purpose of increasing player engagement and the likelihood of players experiencing agency in a system. For instance, many popular video games allow players to make choices within the environment, follow non-linear paths through the game (e.g., Grand Theft Auto V) or select between several available mini-games (e.g., Nintendo Land). Likewise, educational games often include choices for how to progress through a system as a means to support player agency and increase

replayability (Spires, Rowe, Mott, & Lester, 2011; Snow, Jacovina, Allen, & McNamara, 2014). Popular commercial games frequently allow players to customize the visual appearance of game features to their preference (e.g., a player's avatar in World of Warcraft), and this game-based feature (i.e., choice) has been associated with increased immersion and intention to replay a game (Schmierbach, Limperos, & Woolley, 2012; Teng, 2010; Yee, 2006). Similarly, personalization enhances students' motivation and learning outcomes (Cordova & Lepper, 1996).

One of the most effective features that has been incorporated into educational games is user choice. Choices made by individual players have the potential to provide students with a sense of agency, as they are engaging on intellectual or emotional levels and prompt players to persist in their play (Schönau-Fog & Bjørner, 2012). Mechanics and features that promote engagement in these ways are found in educational games such as *Crystal Island* (Lester, Mott, Robison, Rowe, & Shores, 2013) and *Quest Atlantis* (Barab, Pettyjohn, Gresalfi, Volk, & Solomou, 2012), where students are immersed in 3-D game environments. In *Crystal Island*, for example, players control an avatar and explore an island where an illness has recently spread. Players interact with both the environment and other game characters to discover information about this outbreak, and in the process, learn microbiology course content. An important advantage for *Crystal Island* over traditional instruction is that it can promote a strong sense of agency, as students have control over how they obtain knowledge in this environment. A study examining students' performance during the game and on posttest content questions found that students who did well in the game also did well at posttest, and these students were more successful at gathering information during play (Rowe, Shores, Mott, &

Lester, 2010). Students who did not do as well in the game, however, also scored lower at posttest and demonstrated less successful information-gathering behaviors. These findings suggest that when students successfully take agency over their learning experience, they achieve higher outcomes in terms of learning the target material.

Despite the generally positive effects of system choices, research suggests that the inclusion of user control may not be universally beneficial for all students (Katz, Assor, Kanat-Maymon, & Bereby-Meyer, 2006). For instance, while some studies have shown added elements of user choice to be associated with positive outcomes (Cordova & Lepper, 1996; Reynolds & Symons, 2001), others have shown inconsistent (Flowerday & Schraw, 2003), neutral (Parker & Lepper, 1992), or negative effects (Flowerday, Schraw, & Stevens, 2004; Iyengar & Lepper, 2000). One reason for these conflicting results may be that users react differently when presented with increased levels of control. Some students may regulate their behaviors and exert control over the environment—inspiring strong feelings of agency—while others may struggle to set goals and make decisions (Zimmerman, 1990). The ability to exert control during a learning task is challenging for many students, as they often struggle to actively monitor their behaviors (Ellis & Zimmerman, 2001). Overall, research suggests that the inclusion of user control (e.g., choice) has the potential to increase learning outcomes among students; however, these effects may vary based on individual differences in users’ ability to control their behaviors (McNamara & Shapiro, 2005).

Assessment of Student Agency

Students’ inconsistent ability to exert control over their environment has posed an assessment problem for researchers, as it is difficult to capture fine-grained behavior

variations. Traditionally, students' feelings of agency have been assessed through self-reports (Ellis & Zimmerman, 2001; Zimmerman, 1990). These direct measurements are static in nature and often miss out on behavior patterns that emerge over time. An alternative to self-report metrics is the use of stealth assessments (Shute, 2011; Shute, Ventura, Bauer, & Zapata-Rivera, 2009). Stealth assessments are used to covertly measure some attribute or construct without explicitly disrupting the student (Shute, 2011; Shute et al., 2009). These assessments have previously been used to measure a variety of constructs, including students' study habits (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007) and self-regulation ability (Sabourin, Shores, Mott, & Lester, 2012). Game-based systems offer researchers a novel form of stealth assessment through the use of log data (e.g., keystrokes, mouse clicks, choice patterns).

An important goal for researchers is to devise measurements and methods for analyzing the log data that is generated within these game-based systems. Indeed, vital information may emerge by testing the degree to which variations in students' behavior patterns (as indicated by the system log data) shed light on how students experience agency, and whether those experiences influence learning outcomes. One potential approach to measuring individual differences in controlled behavior is through the use of analyses inspired by dynamical systems theory. Dynamical analysis focuses on the complex behaviors that emerge within a given environment; thus, time is treated as a critical variable in addressing patterns of variation and consistency. Because of the focus on time, dynamical methodologies offer scientists a novel means of classifying variations in students' behavior patterns when they are given agency within an adaptive system. Researchers have previously used dynamical methodologies to investigate variations in

behavior patterns within various adaptive systems (Hadwin et al., 2007; Snow, Allen, Russell, & McNamara, 2014; Snow, Jackson, & McNamara, 2014; Snow, Likens, Jackson, & McNamara, 2013; Zhou, 2013). This work has demonstrated the potential for dynamical methodologies to capture nuanced and fine-grained patterns that reveal how students approach learning tasks embedded within adaptive environments.

CURRENT WORK

The current work is comprised of two studies designed to investigate how agency may manifest within system log data and the ultimate impact that agency has on learning outcomes. In Study 1, we employ three novel dynamical methodologies to investigate how variations in behavioral patterns emerge when students have the potential to exert high levels of control (i.e., they have many choices) within an adaptive environment. Although this level of control should lead to increased levels of perceived agency for students, some may struggle to exert control over such an open environment. In Study 1, we examine *how* students interact with the game-based system iSTART-2, and the subsequent learning outcomes associated with those behavior patterns. In particular, we are interested in the impact that controlled and disordered interaction patterns have on target skill acquisition. We hypothesize that students who experience higher levels of agency will exhibit more controlled patterns of behavior and show higher levels of target skill acquisition compared to students who experience lower levels of agency and act in more disordered patterns. Such findings would support previous work showing that when students exert high levels of agency, they are employing controlled and strategic plans of action (Bandura, 2001) and their subsequent learning performance increases (Zimmerman, 1990). If particular interaction patterns (e.g., controlled patterns) are associated with higher levels of performance, our agency-based account predicts that

students will not experience the benefits of those interaction patterns in the absence of choice. Study 2 tests this prediction by removing the opportunity to experience agency by randomly assigning students to an interaction pattern that they must follow. Both Study 1 and Study 2 are presented within the context of the game-based system, iSTART-2.

iSTART-2

The iSTART (Interactive Strategy Training for Active Reading and Thinking) program was designed to provide high school students with instruction on the use of self-explanation and comprehension strategies (Jackson & McNamara, 2013; McNamara, Levinstein, & Boonthum, 2004; McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007). Studies have confirmed that students' comprehension and self-explanation ability is enhanced when they are provided with iSTART strategy instruction (Jackson & McNamara, 2013; McNamara et al., 2007; O'Reilly, Sinclair, & McNamara, 2004; Taylor, O'Reilly, Rowe, & McNamara, 2006). iSTART-2 (Snow, Jacovina et al., 2014) is the most recent version of iSTART. This system provides students with strategy training within the context of a game-based environment. This environment is designed to enhance students' engagement and persistence during prolonged periods of training. When students are engaged and express interest in a learning task, they are more likely to show positive learning outcomes (Pintrich, 2000). The game-based features included in iSTART-2 are designed to enhance students' motivation and engagement during training and, depending on various circumstances, improve learning outcomes (McNamara, Jackson, & Graesser, 2010).

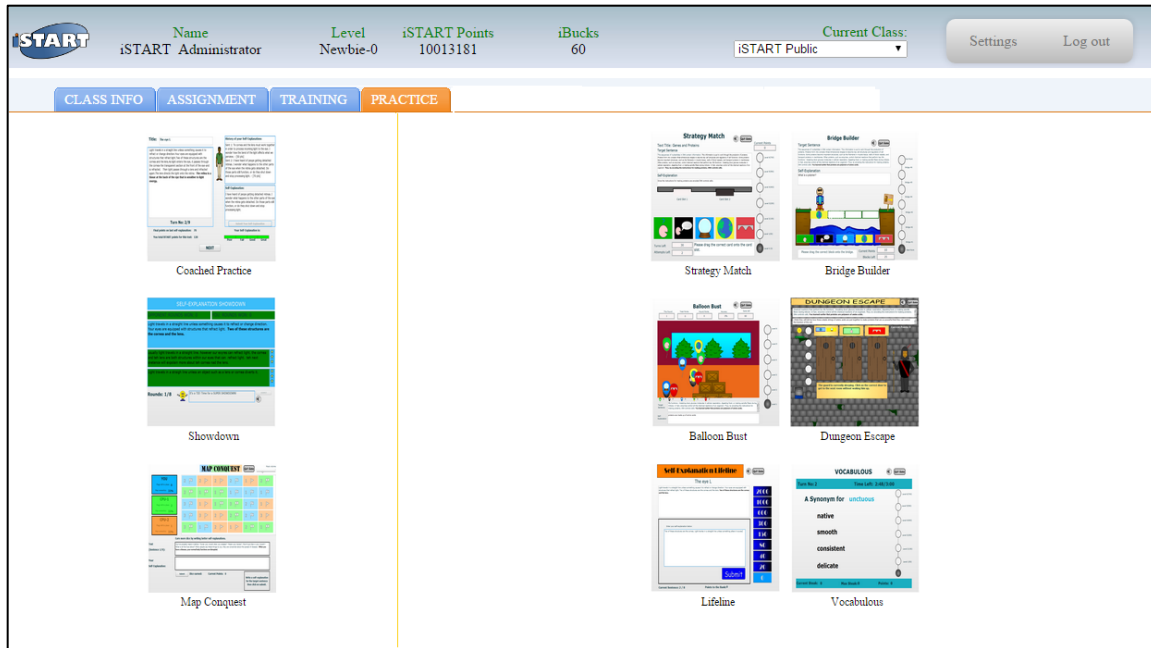


Figure 1. Screen shot of iSTART-2 Selection Menu

iSTART-2 consists of two phases: training and practice. Within the training phase, students are introduced to a pedagogical agent who explains and defines the concept of self-explanation. This agent also discusses the iSTART-2 comprehension strategies: comprehension monitoring, predicting, paraphrasing, elaborating, and bridging. Within this phase, students are provided with examples of each comprehension strategy in separate lesson videos. At the end of each lesson video, students are given a short quiz that assesses their understanding of the strategy. During the practice phase of iSTART-2, students are transitioned to an interactive game-based interface. Within this interface, students can read and self-explain new texts, personalize different aspects of the interface and play mini-games (see Figure 1).

In addition, students can check their personal accomplishments within the system by viewing achievement screens that update students on their current level, number of points earned, and total trophies won. Students increase their level within the system by

earning points when interacting with two different types of generative practice where they write their own self-explanations: Showdown, and Map Conquest. These generative practice games were designed to engage students' interest while they practice using strategies by generating self-explanations. For example, in Showdown, students are asked to generate a self-explanation for numerous target sentences while competing against another player. The student's and computer player's generated self-explanations are compared and the highest quality self-explanation wins the round and any subsequent points (see Figure 2).

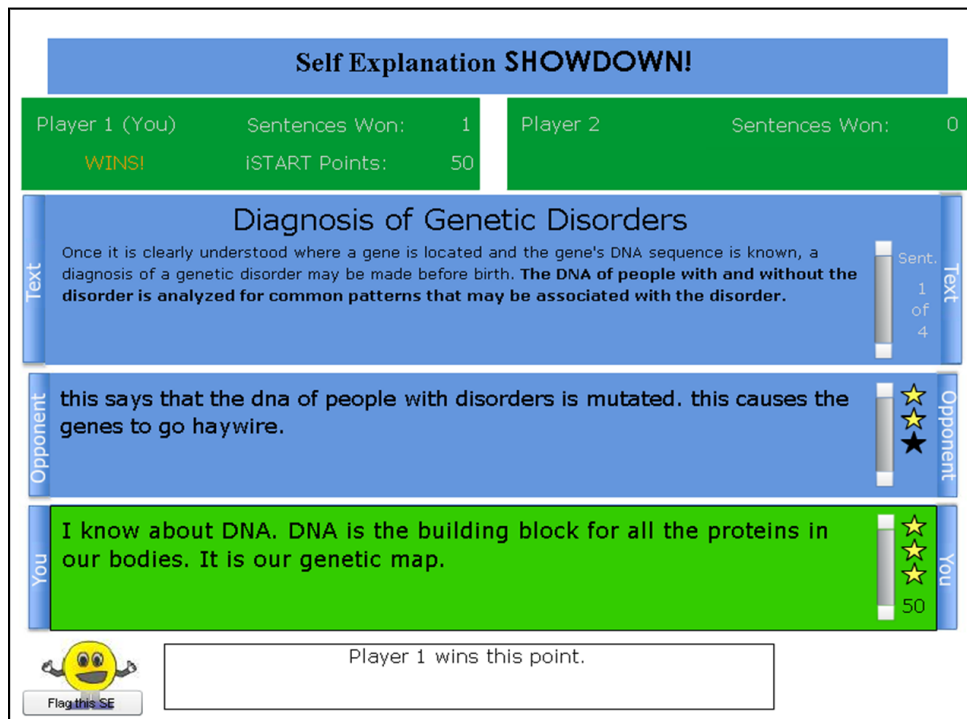


Figure 2. Screen shot of Showdown

As students earn more points within the system, they progress through a series of levels ranging from 0 to 25. Every level progression requires more points to proceed than the previous level, which ensures that students have to exert more effort to progress to higher levels within the system. The points that students earn throughout their

interactions with iSTART-2 also serve as a form of currency (iBucks) that can be used to purchase game-based features. Within iSTART-2, each game-based system feature costs 300 iBucks per interaction. There are two ways students can choose to spend their earned iBucks: personalize the system interface and play mini-games. Students can choose to personalize the system interface by editing an avatar or customizing the background theme. When students choose to edit their avatar, they have the option to change the hairstyles and accessories that their avatar displays in the interface. If students choose to edit their background theme, they can modify the color of the interface (24 total color options). These two features were built into the interface to afford students a sense of control over the environment. However, both of these features were designed to be off-task and tangential to the learning goal of the system.

Students can also choose to spend their earned iBucks on a suite of four mini-games. These mini-games were added to iSTART-2 as a means of identification practice for the self-explanation strategies that the students learned. Each mini-game allows students to engage in play while at the same time practicing reading comprehension strategies. Mini-games are designed to be on-task and act as an extension of the learning goal of iSTART-2. Although these games vary in their game mechanics, the strategy identification task is very similar in each. One example of an iSTART-2 mini-game is Balloon Bust. In Balloon Bust, students are presented with a text and a self-explanation. Students decide which previously learned strategy was used to generate the self-explanation and click on the corresponding balloon to pop it (see Figure 3). In each game, students are given a score based on their performance and a corresponding trophy, if applicable. Students can accumulate trophies throughout their time in the system and

view their performance on the achievement screen tab (i.e., points, levels, and trophies) on the main interface menu.

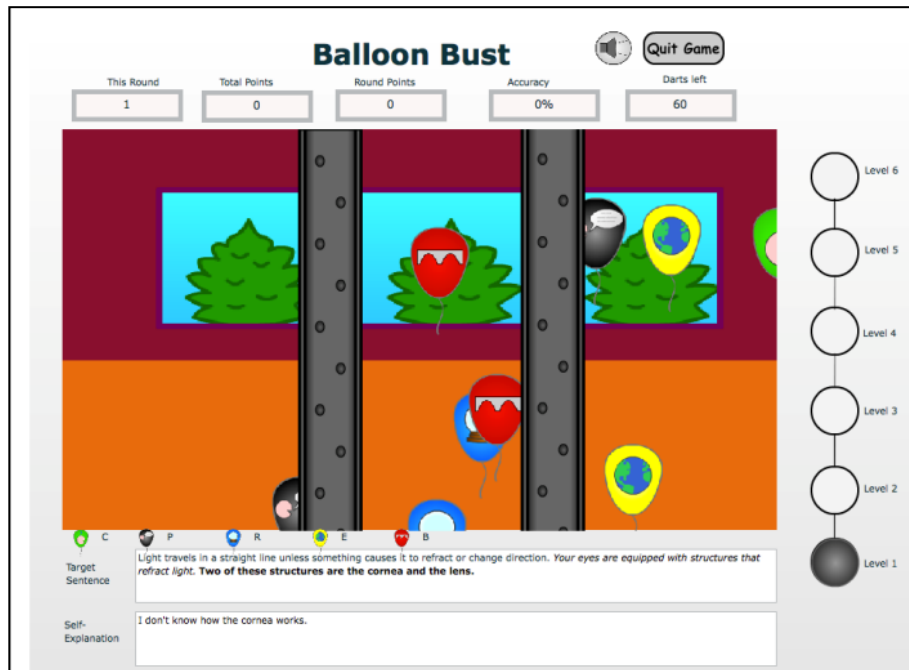


Figure 3. Screen shot of Balloon Bust

Each of these game-based features (i.e., personalizable features and mini-games) is available in the iSTART-2 interface. The system allows features to become unlocked as students progress to higher levels within the system. For instance, the Mohawk hairstyle for avatars can be locked until students reach level 11. The system was designed this way to ensure that some features can act as incentives to enhance students' performance and effort within the system (for more information about the iSTART-2 design, please see Jackson & McNamara, 2013). However, for the current study, all features were unlocked and accessible to the students from the outset.

STUDY 1

Game-based learning environments frequently include elements designed to increase feelings of agency and to provide students with control over their learning trajectory.

Increasing students' agency generally has a positive effect on learning; however, the benefits accrued by students depend on their ability to successfully take control over their environment. Dynamical analyses provide the means to assess students' choice patterns within learning environments and to capture fluid changes in fine-grained behavior patterns. These fine-grained measures potentially afford a deeper understanding of how students vary in their ability to exert agency over their environments and how those behavior patterns relate to learning outcomes. Study 1 uses three dynamical techniques (random walks, Euclidean distances, and Entropy scores) to visualize and quantify students' choice patterns within the iSTART-2 interface. Using these methodologies, we investigate how variations in choice patterns emerge and ultimately impact students' learning outcomes (i.e., self-explanation quality) within the context of iSTART-2.

STUDY 1 METHOD

Participants

This study included 75 college students from a large university campus in Southwest United States. These students were, on average, 18.8 years of age (range: 18-24 years), with a mean reported grade level of college freshman. Of the 75 students, 57% were male, 56% were Caucasian, 23% were Asian, 5% were African-American, 11% were Hispanic, and 5% reported other nationalities.

Procedure

This study included one 3-hour session consisting of a pretest, strategy training (via iSTART-2), extended game-based practice within iSTART-2, and a posttest. At pretest, students were asked to answer a battery of questions that assessed their prior motivation and attitudes. During training, all students watched the iSTART training videos, which

instructed them on the application of self-explanation strategies. After students watched the training videos, they were transferred into the game-based practice menu embedded within iSTART-2. During their time within the game-based practice menu, students were free to interact with the system interface anyway they chose. Students spent approximately 2 hours within the game-based interface. After they finished game-based practice, all students were transitioned to the posttest, where they completed attitude questionnaires similar to those in the pretest.

Measures

Strategy Performance. During game-based practice, students' generated self-explanation quality was measured using an algorithm that combines both Latent Semantic Analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007) and word-based measures (Jackson, Guess, & McNamara, 2010; McNamara et al., 2007). Within iSTART-2, all self-explanations are scored on a scale from 0 to 3. Students are assigned a training self-explanation score by averaging the scores of all their generated self-explanations.

System Interaction Choices. In this study, students were free to interact within the iSTART-2 system. The game-based features fall into one of four types of game-based feature categories. Each type of game-based feature category represents a different functionality within the system interface.

1. *Generative practice games* include game-based practice environments (i.e., Map Conquest, and Showdown) where students generate their own self-explanations. Within each of these practice games students receive feedback regarding their self-

explanations. This provides each student the opportunity to apply and revise comprehension strategies to challenging texts.

2. *Identification mini-games* include four game-based practice environments designed to reinforce learning strategies by asking the students to identify the self-explanation used to generate example self-explanations. These mini-games are designed to provide an alternative form of strategy practice (i.e., strategy recognition) to students.
3. *Personalizable features* (i.e., avatar and background customization) provide students with the opportunity to customize the system interface. These elements were designed to afford students a feeling of personal investment during long-term practice.
4. *Achievement screens* provide students with the opportunity to view their earned trophies, self-explanation scores, levels, and points within the system. These menus were embedded within the system to provide students with the opportunity to monitor their progress during strategy practice.

Game Performance. In both the generative practice and identification mini-games, students earn trophies based on their performance. Students can earn bronze, silver, and gold trophies throughout their time in iSTART-2 and can view these accomplishments on the achievement screen.

Posttest Attitudes. At posttest, students were asked to answer three questions that assessed their confusion, feelings of control, and boredom while engaged within the system (see Table 1). These single-item measures have been used in previous studies to assess individuals' motivation and enjoyment within adaptive systems (e.g., Jackson & McNamara, 2013; Snow, Jackson, et al., 2013).

The use of single-item versus multi-item measurements has been subject to much debate within the psychological literature (De Boer et al., 2004). Indeed, proponents of multi-item measures (i.e., 3 or more items designed to measure one construct) argue that the use of multidimensional scales improves the validity and reliability of the intended measurement (Nunnally & Bernstein, 1994; McIver & Carmines, 1981). However, the use of single-item measures has been shown to have practical advantages. For instance, Robins and colleagues (2001) argued that the use of single-item measures can help reduce the fatigue, frustration, and boredom that is typically associated with high redundancy in multi-item scales. This practicality has led to the wide spread use of single-item measures as a way to assess constructs such as, intelligence (Paulhus, Lysy, & Yik, 1998), agency (Metcalf et al., 2013), and anxiety (Davey, Barratt, Butow, & Deeks, 2007).

In the current study, all single-item measures were presented as forced choice survey scales ranging from 1 to 6 (1 = strongly disagree; 6 = strongly agree; see Table 1). Even-numbered scales remove the opportunity for students to adopt a middle, *neutral* stance (e.g., no opinion, I do not know, not applicable). That is, selecting 1 to 3 indicates some level of disagreement with the statement and selecting 4 to 6 indicates some level of agreement. The use of a neutral option within Likert-scales presents a variety of empirical issues (Moors, 2008). First, when middle or neutral options are explicitly offered, participants can be more likely to choose them (Bishop, 1987). Furthermore, it has been argued that interpretations of middle responses vary aside from true neutrality (e.g., the item is not applicable, or the respondent is unwilling to answer; Stone, 2004). This ambiguity has proven problematic for researchers attempting to establish construct

validity (Klopfer & Madden, 1980; Stone, 2004). One goal of the current work is begin to establish the validity of the relation between self-report measures of agency and actual observed behaviors within an adaptive learning environment. Thus, in this work, we employ a forced-choice design in an attempt to eliminate the ambiguity and validity concerns associated with a neutral option.

Table 1.

Posttest attitude measures

Dependent Measure	Response Statement	Response Scale
Confusion	“I was confused about what I should be doing”	1 - 6
Control	“I felt like I had no control over the system”	1 - 6
Boredom	“I felt bored in the system”	1 - 6

1 (Strongly Disagree) to 6 (Strongly Agree)

Data Processing

During students’ time in the game-based practice menu, all choices (e.g., to play a mini-game or update their avatars) were recorded. These data logs were then used to categorize each interaction as one of the four game-based feature category types (i.e., generative practice games, identification mini-games, personalizable features, and achievement screens).

QUANTITATIVE METHOD

Variations in students’ behavior patterns were assessed using three dynamical methodologies: random walks, Euclidean distances and Entropy analyses. These methodologies afford the opportunity to quantify variations in students’ choice patterns and examine how these different trajectories impacted students’ learning outcomes (i.e.,

self-explanation quality) within the context of iSTART-2. A description and explanation of random walks, Euclidean distances, and Entropy analyses are described below.

Random Walks

Random walks generated a representation of each student's unique interaction trajectory within iSTART-2. This mathematical tool provides a spatial representation of patterns within categorical data as they manifest across time (Benhamou & Bovet, 1989; Lobry, 1996). Each student's trajectory within the system was represented by first examining the sequential order of interactions with various game-based features. Each game-based feature category was assigned an orthogonal vector along an X, Y scatter plot (see Table 2). The assignment of these vector locations is random and not associated with any qualitative value. Random walks have previously been used to trace students' interaction patterns within the game-based system, iSTART-ME and within the writing tutor, Writing Pal (Allen, Snow, & McNamara, 2014; Snow, Allen, Jackson, & McNamara, 2014; Snow, Jacovina, et al., 2014; Snow, Likens, et al., 2013). Each student's random walk began at the origin (0,0). Then, using the system log data to examine the sequential order of students' choices, the particle moves in a manner consistent with the vector assignment of the specified choice. The culmination of the movement of the particle results in a continual trajectory or "walk" that visually represents each student's time within the iSTART-2 system.

Table 2.

Vector assignments

Game-Based Feature	Vector Assignment
Generative Practice Games	-1 on X-axis (move left)
Identification Mini-Games	+1 on Y-axis (move up)
Personalizable Features	+1 on X-Axis (move right)
Achievement Screens	-1 on Y-axis (move down)

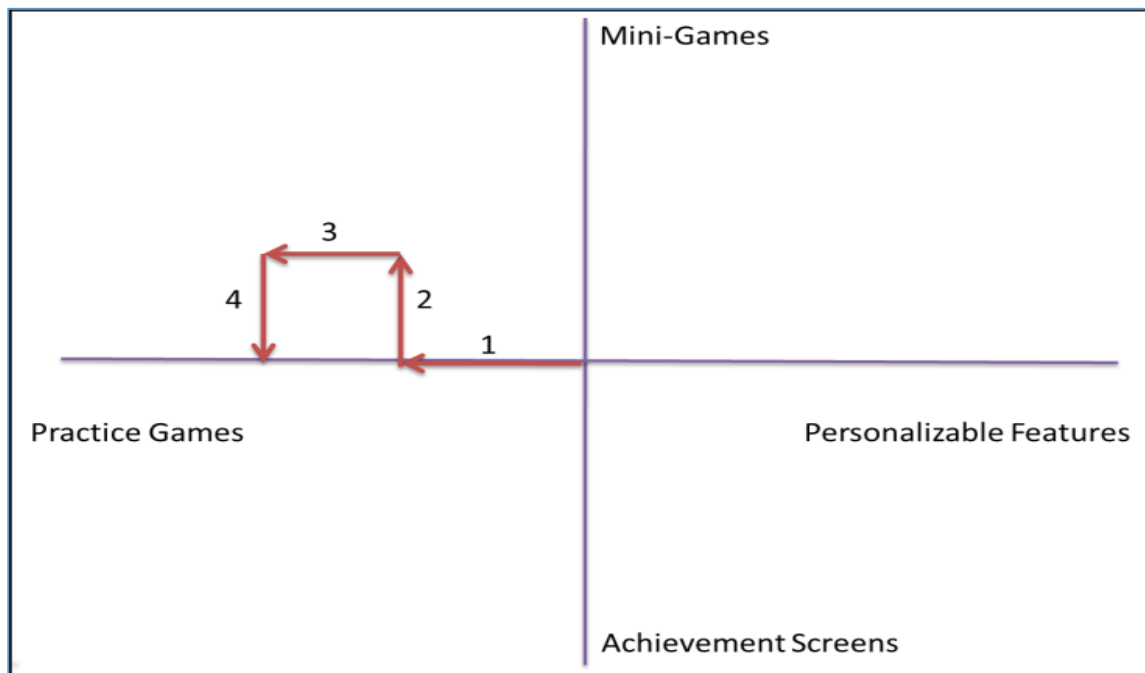


Figure 4. Example random walk

To illustrate what a random walk might look like for a student who made four choices within the iSTART-2 system, see Figure 4. For all students, the starting point of their walk is (0, 0); this is where the horizontal and vertical axes intersect. In this example, the first interaction the student chose was a generative practice game; so, the particle moves one unit to the left along the X-axis (see # 1 in Figure 4). The second interaction choice that the student made was to play a mini-game; thus, the particle moves one unit up along the Y-axis (see # 2 in Figure 4). The student’s third interaction

choice was with another practice game, which moves the particle one unit left along the X-axis (see # 3 in Figure 4). The student’s fourth interaction choice was with an achievement screen, which moves the particle one unit down along the Y-axis (see # 4 in Figure 4). Using these simple rules, a random walk was calculated for every student (n=75) who interacted with iSTART-2.



Figures 5 & 6. Actual Random Walks from Study 1

Figures 5 and 6 are students’ actual walks generated to represent their time spent within iSTART-2. The student represented in Figure 5 reveals a walk with an upward trajectory; we can then infer that this student interacted predominantly with the identification mini-games. Conversely, the student represented in Figure 6 demonstrated a rightward walk trajectory. Thus, we can infer that this student spent the majority of the time interacting with the personalizable features. Interestingly, both walks reveal that students interacted with a multitude of features, as evidenced by the fluctuations in the

figures. Overall the use of random walks provides researchers with a means to visualize trajectories and patterns of choices.

Euclidian Distance

As illustrated in Figure 5, random walks capture fluctuations in students' choice patterns that may inform researchers on the degree to which a student is controlled or deliberate when approaching a learning task. To quantify these fluctuations, distance time series were calculated for each student by using a measure of Euclidean distance. For each student, Euclidian distance was measured from the origin (coordinates 0,0) to each step within his or her walk (see Equation 1). In the Euclidian distance equation, y and x represent the particle's place on the y -axis and x -axis, respectively, and the i represents the i th step within each walk:

$$\text{Distance} = \sqrt{(y_i - y_0)^2 + (x_i - x_0)^2} \quad (1)$$

A Euclidean distance was calculated for each step in a student's walk, which provides a measure of how far the particle moved from the origin. When these distance steps are combined, they produce a distance time series. Distance time series reveal patterns of movement and potentially reveal systematic patterns in learning trajectories through coordinated "steps."

Entropy

Random walks provide a visualization of students' learning trajectories and Euclidian distances help quantify the movements depicted with these walks. In turn, Entropy analyses were conducted to quantify the degree to which these fluctuations are controlled or disordered. Entropy analysis is a statistical measure that has been used previously to measure random, controlled, and ordered processes (Fasolo, Hertwig, Huber, & Ludwig,

2009; Grossman, 1953; Shannon, 1951; Snow, Allen, Russell, et al., 2014; Snow, Jacovina, et al., 2014). In the current study, Shannon Entropy (Shannon, 1951) is used to gain a deeper understanding of how students' choice patterns reflect controlled and ordered processes. To calculate Entropy, we used the distance time series produced from each student's random walk (see Equation 2). Within the Entropy equation, $P(x_i)$ represents the probability of a given interaction. For instance, the Entropy for student X is the inverse of the sum of products calculated by multiplying the probability of each distance by the natural log of the probability of that distance. This formula captures the amount of control and order presented within the distance time series generated within each student's random walk.

$$H(x) = - \sum_{i=0}^N P(x_i) (\log_e P(x_i)) \quad (2)$$

In this study, when a student's choice pattern produces a low Entropy score, it suggests that they demonstrated a highly organized pattern. Conversely, when a student's choice pattern produces a high Entropy score, it suggests that the student demonstrated a disorganized choice pattern. The differences between controlled and disordered behaviors can be visualized using gait patterns. For example, in Figure 7, the footsteps illustrate a systematic and controlled gait. These steps are evenly spaced and ordered. Thus, each step in this pattern comprises a controlled and systematic series analogous to low Entropy. Conversely, Figure 8 reveals a random and disordered gait pattern where the steps are unevenly distributed and unpredictably jump around, illustrative of disorder and thus high Entropy.



Figure 7. Example of controlled pattern (i.e., low Entropy)



Figure 8. Example of disordered pattern (i.e., high Entropy)

Statistical Analyses

To examine the influence of students' behavior patterns within iSTART-2 on their daily performance and posttest attitudes, we conducted Pearson correlation and regression analyses using Entropy scores, average self-explanation quality during training, game performance, and posttest survey responses. In addition, a regression analysis was conducted to examine the degree to which students' Entropy scores accounted for variance in their daily self-explanation quality. Finally, Pearson correlation analyses were conducted to examine the relation between students' Entropy scores and their in-game performance (i.e., trophies won) and also the relation between students' Entropy scores and posttest attitudes.

STUDY 1 RESULTS

Entropy

The current study examined the impact of variations in students' behavior patterns on the in-system performance within iSTART-2. An Entropy analysis was calculated to quantify the fluctuations that manifest within each student's random walk. In the current study, Entropy scores varied considerably, suggesting that students' behavior patterns ranged from controlled to disordered (range=1.32 to 2.32, $M=1.83$, $SD=0.24$; skew =-.22; kurtosis = -1).

Interaction Choices

The relation between Entropy scores and students' frequency of interaction choices was calculated by using Pearson correlations. Results from this analysis revealed no significant relations between students' Entropy scores and the frequency of interactions with generative practice games ($r=-.04$, $p=.73$), identification mini-games ($r=.17$, $p=.13$), personalizable features ($r=-.05$, $p=.66$), or achievement screen views ($r=-.11$, $p=.37$). This suggests that controlled patterns of interactions were not related to any specific feature within iSTART-2.

In-System Performance

Self-Explanation Quality. To examine the effects of controlled interaction patterns on self-explanation quality, a regression analysis was conducted. In this analysis, we examined how Entropy was related to students' average self-explanation quality score from their time within the iSTART-2 system. This analysis revealed a significant relation between students' Entropy scores and their average self-explanation quality scores, $F(1, 74)=4.33$, $p=.041$, $R^2=.06$. These results indicate that students who engaged in more

controlled interaction patterns generated higher quality self-explanations relative to students who engaged in more disordered interaction patterns.

In-Game Performance. Within iSTART-2, students also earned trophies based on their performance within the practice games. To examine the relation between students' patterns of interactions and game-performance, Pearson correlations were conducted. Results from these analyses revealed a significant negative relation between Entropy scores and generative practice trophies ($r=-.26, p=.02$). This means that students with a higher Entropy score (indicative of disordered choice pattern) tended to earn fewer trophies within the system. Again, students who interacted in a more ordered and controlled way showed more success within the system.

Posttest Attitudes

To assess how Entropy related to students' self-reported feelings of confusion, boredom and control, Pearson correlation analyses were conducted. Results from these analyses revealed no significant relation between students' Entropy and their self-reported confusion ($r=.03, p=.78$) or their self-reported boredom ($r=.07, p=.55$). However, Entropy was significantly positively related to students' feeling of lack of control ($r=.26, p=.02$). Thus, students' who engaged in more controlled interaction patterns also reported higher feelings of control. Importantly, the lack of control was not related to other factors such as confusion or boredom.

STUDY 1 DISCUSSION

Enhanced feelings of agency have been found to increase students' engagement and ultimately improve learning outcomes (Flowerday & Schraw, 2000). Learning environments attempt to leverage these findings by adding elements of control for

students. However, students vary in their ability to effectively control and regulate their behaviors when presented with opportunities to exert control (Zimmerman, 1990). Thus, some students may experience high levels of agency and thrive in an environment where they are presented with many choices, whereas others may struggle to effectively exert control over their behaviors and develop a plan of action (McNamara & Shapiro, 2005). Study 1 captured these fluctuations in students' ability to control their behaviors through the use of dynamical methodologies.

Entropy was explored as a means to provide a stealth assessment of students' patterns of interactions within iSTART-2. The use of stealth assessments is important for researchers because they can obviate the need for intrusive and explicit questions (e.g., "How in control are you right now?") which may disrupt students' flow within game-based systems and their subsequent ability to accurately report feelings of agency. In Study 1, we interpreted students with a low Entropy score as interacting with the system with purpose and control. Conversely, when students' choice patterns produced a high Entropy score, they demonstrated a lack of purpose or control. Each student's Entropy score reveals a trend across time, suggestive of the degree to which each student exerted agency within the iSTART-2 system. Our interpretation of the Entropy scores is further supported by students' posttest self-reported feelings of control. Specifically, students who had a more disordered interaction pattern also reported feeling less in control. This relationship between Entropy and feelings of control begins to provide concurrent validity that Entropy may be one way to covertly assess students' feeling of agency over their environment.

In addition to students' attitudes, Entropy scores were also related to students' *performance* within the iSTART-2 system. In particular, when students demonstrated a more controlled interaction pattern (i.e., low Entropy) within the system, they generated higher quality self-explanations and performed better in the practice games during training compared to students who exhibited a disordered interaction pattern (i.e., high Entropy). It is important to note that Entropy was not significantly related to any specific game-based feature. That is, students who tended to behave in more controlled manners did not engage with similar activities within the system interface. This suggests that there are multiple ways to succeed within iSTART-2 and that the impact of students' interactions on learning has less to do with *what* they choose, but *how* they choose to do it. Open interface game-based systems are designed to afford students opportunities to create their own learning trajectory by exploring different activities and features. The results of Study 1 suggest that to be successful within these interfaces, students need to take agency over their learning paths and make meaningful choices. These results are supported by previous work that reveals a positive relation between students' ability to regulate and control their learning behaviors during a task and learning outcomes (Calvert et al., 2005).

An alternative explanation for the results found in Study 1 is that controlled learning trajectories will enhance learning outcomes *regardless of students' agency* in determining those trajectories. This explanation does not rely on students experiencing a sense of agency. According to this hypothesis, as long as students interact with the system in a controlled pattern, they should benefit regardless of their perceived level of agency. If, indeed, agency is of secondary importance, assigning students with a

predetermined interaction pattern should yield similar results. In Study 2, we tested this alternative hypothesis by removing students' agency within the system and assigning them to previously generated interaction patterns. This alternative hypothesis would be supported if students who engage in controlled patterns still outperform students who engage in disordered patterns, despite the patterns being assigned rather than chosen. On the other hand, an agency-based hypothesis would be supported if performance differences do not emerge between students who engage in controlled and disordered interaction patterns.

STUDY 2

Study 1 examined how students' self-selected interaction patterns influenced system performance and perceptions of control. Study 2 builds upon this work by examining whether the positive influence of controlled behavior patterns (i.e., based on Entropy scores) on target skill acquisition (i.e., self-explanation quality) persists when the ability to choose that pattern is removed. A main component of agency is students' ability to take control over their learning path and set their own strategic plan (Zimmerman, 2008). This suggests that if agency contributed to the learning benefits observed in Study 1, those benefits should be attenuated when students have little control over their learning. Study 2 is designed to test that assumption by taking away students' ability to experience agency and instead randomly assigning them a learning trajectory within iSTART-2. This manipulation will begin to tease apart the role of agency and specific interaction patterns on students' learning in the system.

STUDY 2 METHOD

Participants

The current study included 70 college students from a large university campus in the Southwest United States. These students were, on average, 19.8 years of age (range: 18-24 years), with a mean reported grade level of college freshman. Of the 70 students, 64% were male, 57% were Caucasian, 23% were Asian, 4% were African-American, 7% were Hispanic, and 9% reported other nationalities.

Procedure

The procedures for Study 2 were identical to the Study 1 procedures for pretest, strategy training, and posttest. The sole difference occurred when students were transitioned into the game-based practice portion of the experiment. In this section, instead of having free choice over their interaction trajectory, students in Study 2 were randomly assigned an interaction trajectory that was previously identified as controlled or disordered. Students received a list of actions that they were asked to complete sequentially. At the beginning of the practice section of the experiment, an experimenter gave a brief explanation of each of the four types of actions that appeared on the list.

Design

Study 2 included two conditions designed to examine the effect of interaction patterns (controlled or disordered) on target skill acquisition. A median split was conducted on the Entropy scores for participants in Study 1 to create two groups of interaction patterns (i.e., controlled interaction condition and disordered interaction condition). Interaction trajectories generated by students who participated in Study 1 were yoked to students in Study 2. Students were randomly assigned one of these previously generated interaction

patterns and were instructed to replicate the list of behaviors while they engaged within the iSTART-2 environment.

Measures

All of the measures for Study 2 were identical to the measures within Study 1.

Data Processing

All students in Study 2 were assigned an interaction trajectory within the iSTART-2 system. Log data was used to validate that students followed their assigned interaction path. All students who participated in Study 2 demonstrated at least a 90% adherence rate. This reveals that most students followed their assigned interaction path with a high level of accuracy.

Statistical Analyses

To examine differences between controlled and disordered students' in-system performance and posttest attitudes, we conducted ANOVA and Bayesian factor analyses using average self-explanation quality during training, game performance (i.e., trophies won), and posttest survey responses including the between-subjects factor of assigned interaction pattern (i.e., controlled or disordered). In addition, separate one-way between subjects ANOVAs were conducted on students' self-explanation quality during training and total trophies won to assess the extent to which the pre-assigned controlled or disordered patterns influenced performance within iSTART-2. Bayes factor analyses were conducted for each of these one-way ANOVAs to assess the probability of a null hypothesis being accurate over an alternative hypothesis. Additionally, three separate one-way ANOVAs were conducted on students' posttest survey responses to examine students' attitudes regarding the system as a function of whether their assigned

interaction pattern was controlled or disordered. Finally, to compare self-explanation quality between students in Study 1 and Study 2, a two-way ANOVA was conducted using average self-explanation quality and Entropy scores.

STUDY 2 RESULTS

Results from Study 1 revealed that students who engaged in a more controlled pattern of interaction performed better within iSTART-2. To investigate whether these results were related to students' ability to take agency over their learning path, we manipulated students' interaction paths in Study 2. Specifically, all students were assigned a previously generated interaction path from the students in Study 1. A median split was calculated on these paths to distinguish between controlled ($M=1.55$, $SD=.29$) and disordered interaction patterns ($M=2.10$, $SD=.10$). Using this median split, we examined differences in system performance between students assigned controlled and disordered interaction patterns.

In-System Performance

Self-Explanation Quality. Results from Study 1 revealed that when students displayed more controlled interaction patterns, they also generated higher quality self-explanations. In Study 2, we calculated a one-way ANOVA to examine if this trend was similar when students were assigned an interaction pattern. Results from this analysis revealed no significant difference between the self-explanation scores for students in the controlled and disordered interaction pattern groups, $F(1,69)=0.11$, $p=.74$.

This finding supports the null hypothesis. However, traditional statistics are not designed to support the null; thus these results may be overestimated. To test the probability of the extent to which the data supports this null hypothesis over an

alternative hypothesis, a two-sample Bayesian t-test analysis was calculated. Bayesian factor analyses are designed to specifically test how probable the support of a null hypothesis is given the t-value and sample size of the test group. This Bayesian factor analysis was conducted using the web-based platform developed by Rouder and colleagues (2009). For this analysis, the sample sizes of the both conditions (controlled interaction pattern group, $n=35$, and disordered interaction pattern group, $n=35$), along with the t-value of the above analysis ($t=.33$), were used to produce a *JZS* (Jeffreys - Zellner - Siow) Bayes factor (Rouder et al., 2009) of 5.26. This JLZ Bayes factor suggests that there is substantial evidence to support the null hypothesis (i.e., five times more likely) over the alternative.

Game-Performance. A similar one-way ANOVA analysis was conducted to examine whether there was a difference in total trophies won between the two interaction pattern groups. Findings from these analyses revealed that there were no significant differences in total trophies won between the controlled and disordered interaction pattern groups, $F(1,69)=2.25$, $p=.12$. This result suggests that when students were assigned to a controlled interaction pattern, they did not perform more successfully than students assigned to a disordered interaction pattern.

To examine the extent to which the data supports this null hypothesis over the alternative hypothesis, two-sample Bayesian t-test analysis was calculated. Again, sample sizes from both conditions were used (controlled interaction group, $n=35$, and disordered interaction group, $n=35$) along with the t-value from the above analysis ($t=.35$). This analysis revealed a *JZS* Bayes factor of 5.21. This *JZS* Bayes factor suggests that there is

substantial evidence to support this null hypothesis (i.e., five times more likely) compared to the alternative hypothesis.

Posttest Attitudes

Three one-way ANOVAs were calculated to examine differences between the controlled and disordered groups of students' self-reported confusion, boredom, and lack of control. Results from this analysis show a significant difference in students' self-reported confusion at posttest, $F(1,69)=9.63$, $p=.003$. Students who were assigned to controlled interaction patterns reported significantly higher amounts of confusion about what they were doing ($M=3.37$, $SD=1.14$) compared to students who were assigned disordered interaction patterns ($M=2.49$, $SD=1.25$). However, there was no significant difference in self-reported feeling of lack of control between the two interaction pattern groups, $F(1,69)=2.72$, $p=.11$, or self-reported boredom, $F(1, 69)=.73$, $p=.39$.

Self-Explanation Comparison for Study 1 and 2

To examine how the influence of interaction patterns on self-explanation quality was influenced by students' agency, a 2 x 2 ANOVA was conducted including the between-subjects factors of study (Study 1 vs. Study 2) and interaction pattern (controlled vs. disordered). This analysis revealed a marginally significant main effect of study on self-explanation quality, $F(3,136)=3.45$, $p=.065$, with students in Study 1 generating higher quality self-explanations ($M=1.67$, $SD=.54$) than students in Study 2 ($M=1.52$, $SD=.45$). There effect of interaction pattern (controlled vs. disordered) was not significant, $F(3,136)=2.83$, $p=.95$. However, there was a significant interaction between study and interaction pattern, $F(3,136)=4.46$, $p=.036$. Students generated higher quality self-explanations when they self-selected to use a controlled interaction pattern ($M=1.83$,

$SD=.57$) compared to those who self-selected a disordered interaction pattern ($M=1.52$, $SD=.46$) or were assigned to either controlled ($M=1.51$, $SD=.46$) or disordered interaction patterns ($M=1.53$, $SD=.44$; see Figure 9). This finding suggests that the degree of control in students' interaction patterns had a greater impact on performance, with the potential to enhance performance, when students were free to make their own choices within the system.

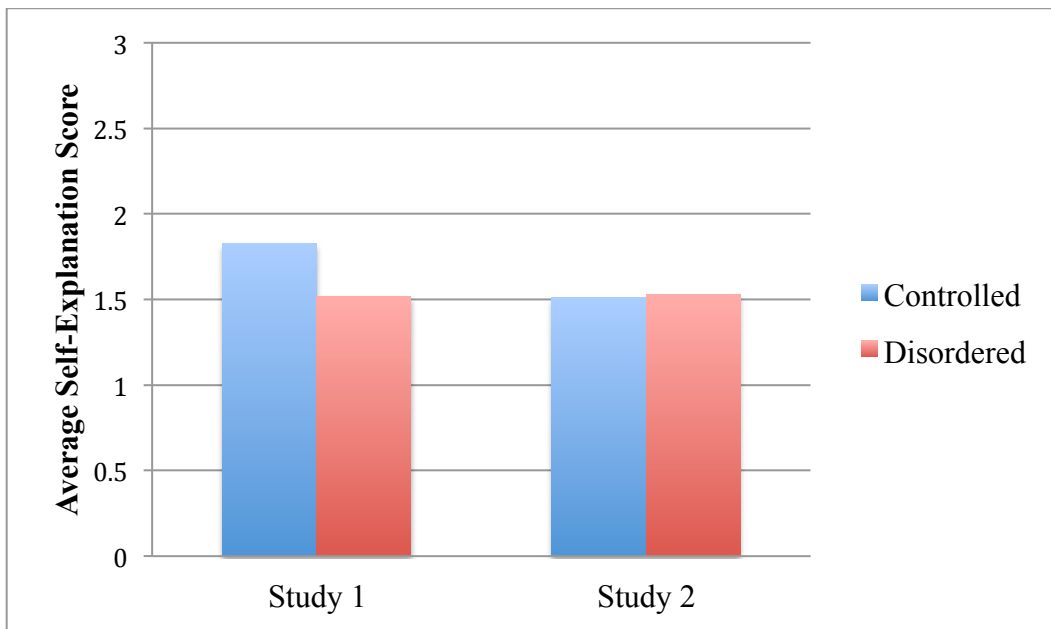


Figure 9. Self-Explanation Scores across Studies 1 and 2

STUDY 2 DISCUSSION

Students who exhibit high levels of agency are said to be engaging in a form of strategic planning (Zimmerman & Schunk, 2001). These students approach a learning task with a goal and subsequently set a plan of action to accomplish this goal. Study 2 was designed to test the impact of agency on learning outcomes in a game-based system. Within iSTART-2, students are afforded high levels of agency and the results from Study 1 revealed that students approached the system in various ways. Some students acted in a

controlled manner (i.e., low Entropy) whereas others acted in a more disordered fashion (i.e., high Entropy). Results showed that when students took agency over their learning path and interacted in a controlled manner within the system, they generated higher quality self-explanations and performed better within the games. Study 2 was designed to tease apart the impact of students' agency and the level of control in their interaction patterns. In particular, we aimed to determine whether the relation of controlled interaction patterns to learning outcomes held when students' agency was removed.

The findings presented here indicate that when students' agency was limited (i.e., they were assigned an interaction pattern), the performance differences between students who engaged in controlled and disorder patterns disappeared. This indicates that students' ability to exert agency and choose their own learning path was related to target performance. Interestingly, in Study 2, students assigned to the controlled condition reported higher levels of confusion than those assigned to the disordered condition. Thus, when agency is inhibited, a controlled pattern may seem redundant and the purpose of such a pattern may be obscured from the student.

This assumption is further supported by the between-experiment analysis on self-explanation quality. This analysis revealed a significant interaction between study (Study 1 vs. Study 2) and interaction pattern (controlled vs. disordered). Specifically, students who engaged in a controlled interaction pattern based on their own choices generated higher quality self-explanations than all other students. This suggests that a controlled pattern only mattered when the student set their own plan and self-selected that interaction path within the system. Thus, assigning students a statistically controlled path without any context may seem just as disorganized to the student as a statically

disorganized path or a self-selected random path. Combined, results from Studies 1 and 2 support our hypothesis that agency is an important component of students' success within adaptive environments.

CONCLUSION

Log data from game-based systems have the strong potential to provide researchers with an opportunity to covertly assess students' ability to exert agency over a learning task (Sabourin et al., 2012; Snow, Allen, Russell, et al., 2014; Snow, Likens, et al., 2013). These systems often allow students to exhibit various levels of control, which influences the interaction patterns that manifest while they interact within the system. This study explored the use of three dynamical methodologies as potential forms of stealth assessment for how students exerted agency within the game-based environment, iSTART-2. Additionally, we examined the impact of these individual interactions on students' students' attitudes and performance. These analyses approach agency in a novel way by examining nuances in students' log data to capture tendencies in their choice selections and behaviors across time. These methodologies may prove useful for the improvement of student models that rely on understanding the relation between students' abilities and performance. Indeed, the tracking and modeling of behavioral trends and patterns over time is critical to our understanding of the various ways in which students exert agency over their environment.

Students vary in their ability to exert agency during learning tasks (Zimmerman, 1990). Results from the current study build upon this work by revealing that students who chose to engage in controlled interaction patterns generated higher quality self-explanations than students who did not. These findings suggest that agency is a key

component of success within a game-based system. Thus, researchers and system designers might strive to find the “sweet spot” of how many choices to include within game-based environments, such that it maximizes the number of students who will experience high levels of agency and minimize the number of students who are overwhelmed and thus respond with disordered choices.

This study builds upon previous work examining how tracing and classifying students’ interactions within adaptive systems can provide information about their ability to regulate and control behaviors (Hadwin et al., 2007; Snow, Jacovina, et al., 2014). The analyses presented here are intended to provide further evidence that dynamical methodologies are valuable tools that can shed light upon various behavioral trends that may manifest within students’ log data. Of course this study is not the end of the story. Future confirmatory studies are needed to further demonstrate the relation between traditional measures of control and regulation (i.e., self-reports) and these statistical techniques. Such work will potentially establish the respective utility of dynamical and traditional measures of learning behaviors (and the subsequent intent behind those behaviors). In Study 1, we found a significant relation between disordered interaction patterns and students’ feelings of lack of control. These results begin to establish the link between agency and controlled interaction patterns. However, one potential weakness that merits further investigation is the use of single-item self-report scales to assess constructs such as control, confusion, and boredom. While previous work has shown that single-item measures can be reliable and potentially increase validity (e.g., Stone, 2004), the affective constructs presented here may indeed be multidimensional and therefore single-item responses may not provide a holistic view of these constructs. Examining

relations between multidimensional self-reports and online choice patterns may further illuminate our understanding of these constructs. Such an approach may afford a more in-depth understanding of how feelings of agency and subsequent behavior patterns emerge overtime.

The ultimate purpose of using dynamical measures is to capture students' behavioral trends they emerge in real time and relate those trends to learning outcomes. Thus, the true measure of the applicability of these measures lies within researchers' ability to implement them in real-time as a means to inform student models. For instance, one critical research question for system developers is how to develop optimal learning trajectories for each individual student. Through the use of visualization and dynamical techniques, systems may be able to recognize non-optimal patterns and steer students toward more effective behaviors. For instance, if students are engaging in disordered behavior patterns, these techniques may be useful in augmenting adaptive environments through the recognition of non-optimal patterns and subsequently prompting students toward a more controlled behavior trajectory.

In conclusion, students' ability to exert agency over a learning task and act in a decisive manner has been shown to be a critical skill for academic success (Bandura, 1989; Hadwin et al., 2007). The current study builds upon this work by revealing that agency is a key component of success within game-based systems. The analyses presented here are among the first attempts to covertly examine how variations in students' interaction patterns (both self-selected and assigned) influence target skill performance. Game-based systems frequently incorporate multiple instructional choices and learning trajectories that afford students the potential for high levels of agency. We

expect that the findings presented here will help shed light upon the impact that interface design can have on students' performance. Overall, these findings support the notion that in order to be successful within game-based systems students need to take agency over their learning paths and make meaningful choices.

CHAPTER 1 REFERENCES

- Allen, L. K., Snow, E.L., & McNamara, D. S. (2014). The long and winding road: Investigating the differential writing patterns of high and low skilled writers. In J. Stamper, S. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 304-307). London, UK.
- Bandura, A. (1989). Human agency in social cognitive theory. *American psychologist*, 44(9), 1175 – 1184.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1), 1-26.
- Barab, S., Pettyjohn, P., Gresalfi, M., Volk, C., & Solomou, M. (2012). Game-based curriculum and transformational play: Designing to meaningfully positioning person, content, and context. *Computers & Education*, 58(1), 518-533.
- Benhamou, S., & Bovet, P. (1989). How animals use their environment: A new look at kinesis. *Animal Behavior*, 38, 375-383.
- Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51(2), 220-232.
- Calvert, S., Strong, B., & Gallagher, L. (2005). Control as an engagement feature of young children's attention to and learning of computer content. *The American Behavioral Scientist*, 48, 578-589.
- Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2009). Combining shared control with variability over surface features: Effects on transfer test performance and task involvement. *Computers in Human Behavior*, 25, 290–298.
- Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2011). Learner-controlled selection of tasks with different surface and structural features: Effects on transfer and efficiency. *Computers in Human Behavior*, 27, 76–81.
- Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88, 715– 730.
- Davey, H. M., Barratt, A. L., Butow, P. N., & Deeks, J. J. (2007). A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. *Journal of clinical epidemiology*, 60, 356-360.

- De Boer, A. G. E. M., Van Lanschot, J. J. B., Stalmeier, P. F. M., Van Sandick, J. W., Hulscher, J. B. F., De Haes, J. C. J. M., & Sprangers, M. A. G. (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life?. *Quality of Life Research*, *13*(2), 311-320.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press, New York.
- Ellis, D., & Zimmerman, B. J. (2001). Enhancing self-monitoring during self-regulated learning of speech. In H. J. Hartman (Ed.), *Metacognition in learning and instruction. Theory, research, and practice* (pp. 205–228). Dordrecht: Kluwer Academic Press.
- Fasolo, B., Hertwig, R., Huber, M., & Ludwig, M. (2009). Size, entropy, and density: What is the difference that makes the difference between small and large real-world assortments? *Psychology & Marketing*, *26*, 254-279.
- Flowerday, T., & Schraw, G. (2000). Teachers' beliefs about instructional choice: A phenomenological study. *Journal of Educational Psychology*, *92*, 634–645.
- Flowerday, T., & Schraw, G. (2003). Effect of choice on cognitive and affective engagement. *Journal of Educational Research*, *96*, 207–215.
- Flowerday, T., Schraw, G., & Stevens, J. (2004). The role of choice and interest in reader engagement. *Journal of Experimental Education*, *72*, 93–114.
- Ford, M. E. (1992). *Motivating humans: Goals, emotions, and personal agency beliefs*. Newbury Park, CA: Sage Publications.
- Grossman, E. R. F. W. (1953). Entropy and choice time: The effect of frequency unbalance on choice-response. *Quarterly Journal of Experimental Psychology*, *5*(2), 41-51.
- Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*, 107-124.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*, 111-127.
- Iyengar, S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, *79*, 995–1006.

- Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology, 105*, 1036-1049.
- Jackson, G. T., Guess, R. H., & McNamara, D. S. (2009). Assessing cognitively complex strategy use in an untrained domain. *Topics in Cognitive Science, 2*, 127-137.
- Katz, I., Assor, A., Kanat-Maymon, Y., & Bereby-Meyer, Y. (2006). Interest as a motivational resource: Feedback and gender matter, but interest makes the difference. *Social Psychology of Education, 9*, 27– 42.
- Klopper, F. J., & Madden, T. M. (1980). The Middlemost Choice on Attitude Items: Ambivalence, Neutrality, or Uncertainty?. *Personality and Social Psychology Bulletin, 6*(1), 97-101.
- Landauer, T. K., McNamara, D. S., Dennis, S. E., & Kintsch, W. E. (2007). *Handbook of latent semantic analysis*. Lawrence Erlbaum Associates Publishers.
- Lester, J. C., Mott, B. W., Robison, J. L., Rowe, J. P., & Shores, L. R. (2013). Supporting self-regulated science learning in narrative-centered learning environments. *In International handbook of metacognition and learning technologies* (pp. 471-483). Springer New York.
- Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biological Evolution, 13*, 660–665.
- McIver, J. P., & Carmines, E. G. (1981). *Unidimensional scaling*. Thousand Oaks, CA: Sage.
- McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007). Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 227-241). Mahwah, NJ: Erlbaum.
- McNamara, D. S., Jackson, G. T., & Graesser, A. C. (2010). Intelligent tutoring and games (ITaG). In Y. K. Baek (Ed.), *Gaming for classroom-based learning: Digital role-playing as a motivator of study* (pp. 44-65). Hershey, PA: IGI Global.
- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers, 36*, 222-233.
- McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I. B. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 397-420). Mahwah, NJ: Erlbaum.

- McNamara, D. S., & Shapiro, A. M. (2005). Multimedia and hypermedia solutions for promoting metacognitive engagement, coherence, and learning. *Journal of Educational Computing Research*, 33, 1-29.
- Metcalf, J., Eich, T. S., & Miele, D. B. (2013). Metacognition of agency: Proximal action and distal outcome. *Experimental Brain Research*, 229, 485–96.
- Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & Quantity*, 42(6), 779-794.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory (3)* McGraw-Hill. New York.
- O'Reilly, T.P., Sinclair, G.P., & McNamara, D.S. (2004). Reading strategy training: Automated versus live. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1059-1064). Mahwah, NJ: Erlbaum.
- Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology*, 62, 625–633.
- Paulhus, D. L., Lysy, D. C., & Yik, M. S. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests?. *Journal of Personality*, 66, 525-554.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychological Review*, 18, 315-341.
- 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, 531-549.
- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology*, 25, 92-104.
- Reynolds, P. L., & Symons, S. (2001). Motivational variables and children's text search. *Journal of Educational Psychology*, 93, 14-22.
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and social psychology bulletin*, 27, 151-161.

- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*, *16* (2), 225-237.
- Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2010). Individual differences in gameplay and learning: A narrative-centered learning perspective. *In Proceedings of the Fifth International Conference on the Foundations of Digital Games* (pp. 171-178). ACM.
- Sabourin, J., Shores, L. R., Mott, B. W., & Lester, J. C. (2012). Predicting student self-regulation strategies in game-based learning environments. *In Intelligent Tutoring Systems* (pp. 141-150). Springer Berlin Heidelberg.
- Schmierbach, M., Limperos, A. M., & Woolley, J. K. (2012). Feeling the need for (personalized) speed: How natural controls and customization contribute to enjoyment of a racing game through enhanced immersion. *Cyberpsychology, Behavior, and Social Networking*, *15*(7), 364-369.
- Schønau-Fog, H., & Bjørner, T. (2012). “Sure, I would like to continue”: A method for mapping the experience of engagement in video games. *Bulletin of Science, Technology & Society*, *32*, 405–412.
- Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, *30*, 50–64.
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer Games and Instruction*, *55*, 503-524.
- Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. *In Serious games: Mechanisms and effects*. (pp. 295-321). Philadelphia, PA: Routledge/LEA, 2.
- Spires, H. A., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). Problem solving and game-based learning: Effects of middle grade students' hypothesis testing strategies on learning outcomes. *Journal of Educational Computing Research*, *44*(4), 453-472.
- Snow, E. L., Allen, L. K., Jackson, G. T., & McNamara, D. S. (2014). Tracking choices: Computational analysis of learning trajectories. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 316-319). Heidelberg, Berlin, Germany: Springer.
- Snow, E. L., Allen L. K., Russell, D. G., & McNamara, D. S. (2014). Who’s in control?: Categorizing nuanced patterns of behaviors within a game-based intelligent

- tutoring system. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 185-192). Heidelberg, Berlin, Germany: Springer.
- Snow, E. L., Jackson, G. T., & McNamara, D. S. (2014). Emergent behaviors in computer-based learning environments: Computational signals of catching up. *Computers in Human Behavior*, *41*, 62-70.
- Snow, E. L., Jacovina, M. E., Allen, L. K., Dai, J., & McNamara, D. S. (2014). Entropy: A stealth assessment of agency in learning environments. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 241-244). Heidelberg, Berlin, Germany: Springer.
- Snow, E. L., Jackson, G. T., Varner, L. K., & McNamara, D. S. (2013). Investigating the effects of off-task personalization on system performance and attitudes within a game-based environment. In S. K. D'Mello, R. A. Calvo, and A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 272-275). Heidelberg, Berlin, Germany: Springer.
- Snow, E. L., Likens, A., Jackson, G. T., & McNamara, D. S. (2013). Students' walk through tutoring: Using a random walk analysis to profile students. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 276-279). Heidelberg, Berlin, Germany: Springer.
- Stone, M. H. (2004). Substantive scale construction. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp.201–225). Maple Grove, MN: JAM Press.
- Tabbers, H. K., & de Koeijer, B. (2010). Learner control in animated multimedia instructions. *Instructional Science*, *38*, 441–453.
- Taylor, R. S., O'Reilly, T., Rowe, M., & McNamara, D. S. (2006). Improving understanding of science texts: iSTART strategy training vs. web design control task. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2234-2239). Mahwah, NJ: Erlbaum.
- Teng, C. I. (2010). Customization, immersion satisfaction, and online gamer loyalty. *Computers in Human Behavior*, *26*, 1547–1554.
- Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, *9*(6), 772-775.

- Zhou, M. (2013). Using traces to investigate self-regulatory activities: A study of self-regulation and achievement goal profiles in the context of web search for academic tasks. *Journal of Cognitive Education and Psychology, 12*, 287-305.
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*, 3-17.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45(1)*, 166-183.
- Zimmerman, B. J., & Schunk, D. H. (2001). Reflections on theories of self-regulated learning and academic achievement. Self-regulated Learning and Academic Achievement: *Theoretical Perspectives, 2*, 289-307.

CHAPTER 2

“Promoting Metacognition within a Game-Based Strategy Tutor”

Manuscript submitted to the Journal of Computers and Education

Every day, people are exposed to a vast amount of information concerning their surrounding environment from television, radio, Internet, and social media outlets. In order to sift through this overwhelming amount of information, people often reflect upon their understanding of the context in which they encounter this information and their prior knowledge of the subject matter. People’s ability to reflect upon what they do and do not know is often referred to as metacognition (Flavell, 1979).

Broadly, the term metacognition refers to the process of thinking about one’s own cognition (Hacker, Dunlosky, & Graesser, 1998). Specifically, this higher-level process relies on two primary sub-components: the *knowledge* and *regulation* of cognition (Brown, 1978; Flavell, 1979; Schraw & Moshman, 1995). Metacognitive knowledge refers to the information that people know about their cognition and cognitive processes, including knowledge about skills and the task criteria, as well as knowledge about the specific strategies that are most effective in different learning situations. Regulating one’s own cognition, on the other hand, typically involves planning, monitoring, and evaluation processes (Brown & DeLoache, 1978; Schraw & Moshman, 1995; Veenman & Spaans, 2005). Thus, in order to successfully regulate behavior, individuals must be able to set goals and select appropriate strategies for the learning task, test their own skills, and re-evaluate their goals and strategies based on their performance.

The development of strong metacognitive skills is crucial for success on a number of academic tasks (Pintrich, 2000; Wang, Haertel, & Walberg, 1990). In order to be

successful and autonomous learners, students need to be able to detect when they have succeeded or failed on a learning task, as well as be able to revise their own learning processes accordingly. Importantly, metacognitive processes aid students in the creation of *new* knowledge about the learning task, as monitoring and evaluative processes can lead them to notice critical deficits in their knowledge (Benjamin, 2003; Pintrich, 2002). Indeed, one of the critical aspects of enacting successful metacognition (i.e., metacognition that leads to learning) relates to the process of accurately assessing one's own performance on a particular task (Andrade & Du, 2007; Falchikov & Boud, 1989). However, like many regulatory skills, students often struggle to accurately assess their own knowledge and appropriately control their cognitive activities (Pintrich, 2000; Varner, Roscoe, & McNamara, 2013; Zimmerman, 2008; Zimmerman & Schunk, 2001).

This discrepancy has led researchers to attempt to identify effective techniques that stimulate and support metacognitive processes during learning tasks. Specifically, within adaptive learning environments, researchers have developed features that provide both *direct* and *indirect* forms of metacognitive support (Graesser & McNamara, 2010). Direct support features provide instruction and feedback about successful metacognitive strategies and skills. System features that provide direct support will often present examples of how the strategies can be implemented and explicitly prompt students to apply these strategies during their studies. For example, MetaTutor is one system that provides students with direct support for metacognition (Azevedo, Johnson, Chauncey, & Burkett, 2011). In MetaTutor, students receive instruction on self-regulated learning processes from animated agents, and they engage in activities that require deep understanding of these processes. Students who engage in these activities prior to

completing a learning task are more likely to show evidence of metacognitive processes such as monitoring (Azevedo et al., 2009). Another example of direct support for metacognition can be found in the Help Tutor (Roll, Alevan, McLaren, & Koedinger, 2011). This system does not provide instructional lessons in metacognition, but provides feedback messages that explicitly encourage helpful metacognitive behaviors. In one study, the Help Tutor was integrated into a geometry tutor to provide students with feedback on their help-seeking behavior (Roll, Alevan, McLaren, & Koedinger, 2011). For example, if a student rapidly requested hints until receiving a bottom-out hint, the system would provide a message that encouraged the students to proceed through the problem more slowly and to read each hint carefully. These feedback messages thus acted as direct support for positive metacognition. Students who received these messages exhibited improved help-seeking behaviors compared to students in a control condition who did not see this metacognitive feedback.

Indirect support features involve scaffolds, prompts, and feedback that encourage learners to monitor their understanding and engage in planning and control processes when necessary, but without providing extensive training on metacognitive strategies. Indirect support features are hypothesized to be particularly helpful for students who already have sufficient metacognitive skills but who may not engage them spontaneously (Bannert, Hildenbrand, & Mengelkamp, 2009). Indirect support features can come in many forms; with a broad definition, any system that provides feedback might be considered to be encouraging metacognition. That is, feedback can prompt students to reevaluate the accuracy of their knowledge or spur unsuccessful students to use a new strategy. However, to avoid such a broad definition of indirect support, we focus on

features that are particularly well suited for supporting metacognition, though indirectly. For example, AutoTutor engages students in a dialogue in which an agent indirectly promotes metacognition through hints, prompts, feedback, and comprehension questions (Graesser & McNamara, 2010). This exchange between the agent and the learner results in deeper comprehension of course content (Nye, Graesser, & Hu, 2014). Although AutoTutor delivers more than metacognitive feedback, its design is informed by theories of metacognition and is therefore mindful of the importance that metacognition has for successful learning. Another example of indirect support for metacognition comes from research conducted to support the development of the MetaHistoReasoning Tool (Poitras, Lajoie, & Hong, 2012). Poitras and colleagues created a learning environment in which students read historical texts and were prompted to answer questions that required them to generate inferences to explain gaps in the text. The goal of these elaborative interrogations was to encourage students to recognize coherence breaks in the text and subsequently fill them in by elaborating. Thus, students who had lapsed in their monitoring while reading would still have an opportunity to overcome the knowledge gap. Students using this tool outperformed students in a control condition on a recall task, and showed more evidence of comprehension monitoring.

Systems, of course, are not limited to either direct or indirect support. Both MetaTutor (Azevedo, 2005) and ANDES (Roll, Alevan, McLaren, & Koedinger, 2011), for example, provide both direct and indirect support. However, balancing the combination of direct and indirect metacognitive support features remains an important research topic. Currently, researchers are striving to promote metacognition without overburdening the learner with too many simultaneous tasks or feedback messages. In

that vein, the current paper builds upon previous work by presenting a series of experiments that test a new feedback and remediation feature designed to provide additional *indirect* metacognitive support within the Interactive Strategy Training for Active Reading and Thinking Tutor (iSTART-2; Snow, Allen, Jacovina, & McNamara, 2015). The goal of the work presented here is to test the utility of adding an additional metacognitive support feature to a system that has previously been shown to improve metacognitive knowledge through direct metacognitive strategy instruction (Jackson & McNamara, 2013).

iSTART-2

The Interactive Strategy Training for Active Reading and Thinking-2 (iSTART-2) is a game-based intelligent tutoring system (ITS) designed to provide high school students with instruction on reading comprehension strategies, specifically focusing on science texts (Jackson & McNamara, 2013; Snow, Allen, Jacovina, & McNamara, 2015; Snow, Jackson, & McNamara, 2014). These strategies are presented during the training phase of iSTART-2 through a series of lesson videos presented by a pedagogical agent. Specifically, iSTART-2 teaches five self-explanation reading strategies (comprehension monitoring, paraphrasing, prediction, elaboration, and bridging). For example, in the comprehension monitoring lesson, students are encouraged to monitor their understanding and self-explain about what does and does not make sense in the text. Similarly, in the elaboration lesson, students learn to self-explain using their prior knowledge to expand on information that is in the text. In each video, the pedagogical agent models the use of the strategy through examples. Modeling in this way can help students who are beginning to internalize new strategies, and can eventually lead to

successful self-regulation in which the strategies are used adaptively (Schunk & Zimmerman, 2007). During the practice phase of iSTART-2, students are able to access a suite of practice games, customize the appearance of the interface, and monitor their recent performance through achievement screens (see Figure 1 for screenshots). In the full version of the system, students can engage in practice within the context of both identification games (i.e., identify a modeled strategy) and generative games (i.e., type their own self-explanation). The current study targeted aspects of the generative games (i.e., Map Conquest, Show Down, and Coached Practice); therefore, those environments will be the focus for the remainder of the article.



Figures. 1 & 2. The game-based practice interface in iSTART-2 and a screen shot of the Map Conquest game.

During the generative games, students read a text and type self-explanations for a number of target sentences. In addition to this practice task, each of these tasks includes certain game elements, such as points or competition. In Map Conquest, for example, students earn a number of flags, based on their self-explanation quality, and these flags can be used to capture enemy territories (see Figure 2). Within all of these games, the quality of students' self-explanations is scored using a linguistic algorithm (scores range from 0 to 3). This scoring algorithm uses a combination of Latent Semantic Analysis

(LSA; Landauer, McNamara, Dennis, & Kintsch, 2007) and word-based measures. A higher score indicates that a student's self-explanation includes information from previous parts of the text and/or outside information that is related to text content. A lower score indicates that a student has not written enough, information was simply repeated or paraphrased, or that irrelevant information was included. The scoring is thus intended to capture the degree to which students integrate new text information with previous information and their prior knowledge of the text topic. Previous work has shown that the algorithm performance is comparable to human raters (McNamara, Boonthum, Levinstein, & Millis, 2007).

CURRENT STUDY

Indirectly Promoting Metacognition within iSTART-2

The lesson videos in iSTART-2 act as direct support for metacognition by encouraging students to be aware of their understanding and by providing explicit strategies that can guide cognitive activities during reading. Once students have completed the lesson videos, however, they are provided with little additional *explicit* support for these metacognitive skills within game-based practice. For instance, in Map Conquest and Showdown, students receive scores on self-explanation quality, but do not receive explicit feedback indicating that they are performing sub-optimally, nor are they prescribed specific strategies or different system activities that can help them to overcome their struggles. One exception is Coached Practice, in which students receive scores and are also asked to indicate which strategy they used while writing their self-explanation, potentially prompting additional feedback from the pedagogical agent. For example, if a student receives a low score on a self-explanation and indicates that only

the monitoring strategy was used, the pedagogical agent might suggest that this student also attempt to relate the target sentence to other ideas from the text. In this way, students are guided to use strategies that allow them to better control their cognitive activities; thus, rather than simply monitoring their understanding, students are encouraged to bolster that understanding by using various comprehension strategies.

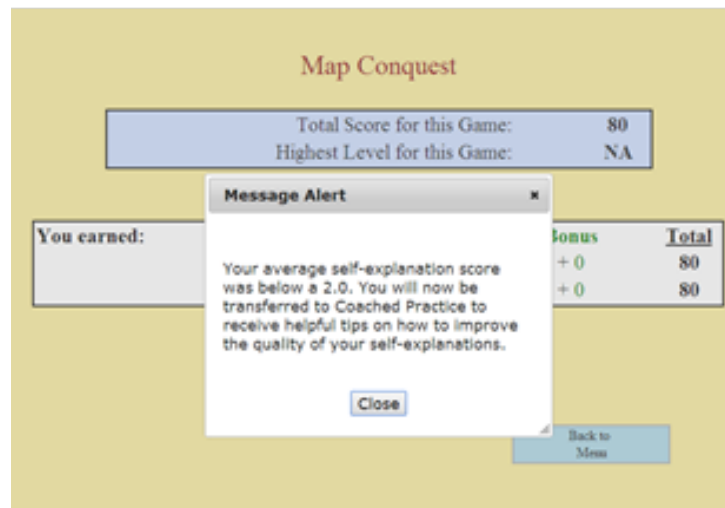


Figure. 3. Screen shot of the pop-up message that appears after students fail to score an average of 2.0 on their self-explanations. The message explains that the student will be transferred to Coached Practice.

In this paper, we present a series of experiments that test a new system feature designed to indirectly enhance metacognition during the game-based practice portion of iSTART-2. This feature explicitly informs students when their performance within Map Conquest and Showdown is low (promoting metacognition) and subsequently transitions them to a remedial Coached Practice environment, where they can receive direct strategy instruction. This *performance threshold feature* is embedded within the practice games of iSTART-2 (Map Conquest and Showdown) and calculates students' average self-

explanation score after each game play. This score is then compared to an experimenter-set threshold, and if this threshold is not met, students are presented with a pop-up message (see Figure 3). The message alerts students that they scored below the threshold (or task standard) and that they will be transitioned to Coached Practice where they receive feedback on how to improve their performance. After closing the pop-up message, students are automatically sent to Coached Practice (see Figure 4). This *performance threshold feature* is designed to provide indirect support for metacognitive monitoring of reading comprehension by presenting students with information on their performance. Specifically, this feature alerts students to their self-explanation score and provides them with a "standard" to which they can compare and attempt to achieve. This comparison is designed to promote self-reflection on how the student's score aligns with the desirable score specified by the system. It is hypothesized that when students are alerted to the misalignment between their self-explanation performance and the systems' desired score (i.e., the experimenter set threshold), they will be more likely to reflect on and consider how they can improve their performance within the system.

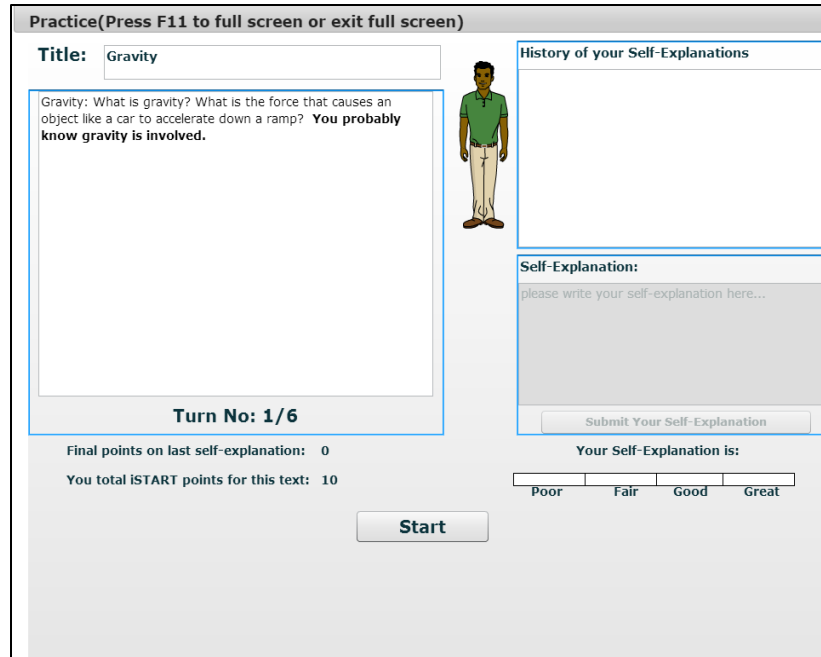


Figure. 4. Coached Practice, the generative activity to which students were transferred after scoring below a 2.0 average during a practice game.

One important aspect of the performance threshold feature is the particular score threshold that is set by the experimenter. If this threshold is not met, students receive a message and are then transitioned to Coached Practice. In the first two studies, we set this self-explanation score threshold to 2.0 (the iSTART algorithm ranges from 0 to 3). When students' self-explanations score above the threshold, it is indicative that they are likely integrating prior knowledge into their responses (Jackson & McNamara, 2011). Previous research has shown that students' use of prior knowledge is critical for their ability to generate and comprehend texts (McNamara & Kintsch, 1996), making the 2.0 threshold an appropriate gauge of success.

STUDY 1

iSTART-2 provides students with direct instruction on metacognitive strategies during the lesson videos. However, within the practice games (Showdown and Map Conquest), there is little to no indirect support for metacognitive strategies or awareness. Thus, a performance threshold feature was created within these games to alert students when their performance is low and then transition them to Coached Practice where they receive feedback on the quality of their self-explanation strategies. Although it is hypothesized that this indirect feedback would benefit students, not all individuals share the same prior knowledge. Therefore it is important to examine how this new feature may impact certain groups of students. Thus, using the new performance threshold feature, we attempted to answer three research questions:

- 1) Does the performance threshold feature (embedded within iSTART-2) accurately use the online data from students' game plays to present messages and transition students to appropriate practice activities?
- 2) How does the transition functionality impact students' self-explanation quality?
- 3) Do the effects of the performance threshold feature vary as a function of students' prior domain knowledge?

STUDY 1 METHOD

Participants

The current work took place on a large university campus in the Southwest United States and included 28 college students who participated for course credit. Students were, on

average, 19.6 years of age (range 18 to 24), 50% were male, 41% were Caucasian, 41% were Asian, 6% were Hispanic, and 12% reported other ethnicities.

Procedure

Study 1 was a 3-hour single-session laboratory study, which consisted of a pretest, strategy training, game-based practice within iSTART-2, and a posttest. During pretest, students answered a battery of questions that assessed their prior science knowledge and motivation toward the learning task. After completing the pretest, students watched the iSTART lesson videos and completed one round of Coached Practice where they were provided with formative feedback on their generated self-explanations. After students completed Coached Practice, they were transferred into the game-based practice menu within iSTART-2. During game-based practice, which included Map Conquest and Showdown (but not Coached Practice), students were free to interact within the system for approximately 1 hour. During this time, their average self-explanation score was calculated for each game and if this average was below a 2.0, they received a message from the system (see Figure 3) and were transitioned to the remedial Coached Practice environment. In this study, students could choose to close out of the remedial Coached Practice environment without completing it. Thus, the message and performance threshold feature served as a *recommendation* that students engage with the remedial Coached Practice environment. Finally, after students completed the game-based practice portion of the experiment, they were transferred to the posttest, where they completed measures similar to the pretest.

Measures

Prior Science Knowledge. Students' prior science knowledge was assessed using a 9-item, four alternative multiple-choice test addressing knowledge of areas such as biology (e.g., "Which of the following tissues produces voluntary body movements?"), and chemistry (e.g., "Which of these has a positive charge and is found in the nucleus of an atom?"). These questions were designed to assess students' general science knowledge and are modified from a previously validated prior knowledge measure (Cronbach's $\alpha = .74$; O'Reilly, Best, & McNamara, 2004).

In-Game Strategy Performance. Students' generated self-explanation quality was assessed while they engaged within the generative practice environments. The previously described iSTART algorithm scored each self-explanation on a scale that ranges from 0 to 3. Thus, students received an average self-explanation score for each generative practice activity (i.e., Showdown, Map Conquest, and Coached Practice) that they completed.

STUDY 1 RESULTS

Games Played

Students were free to interact with the game-based interface for approximately 1 hour. During this time, they could choose to play either Showdown or Map Conquest as many times as they wanted and in whatever order they wanted. On average, students completed 3.96 total game plays (range = 2 to 9, $SD = 1.86$), with Map Conquest being played an average of 2.1 times (range = 1 to 6, $SD = 1.35$) and Showdown being played an average

of 1.8 times (range = 1 to 5, $SD = 1.10$). Although Map Conquest was chosen at a higher rate than Showdown, this difference was not significant ($t(27) = -1.06, p = .30$).

Transitions

A preliminary analysis was conducted to verify the functionality of the performance threshold feature. To conduct this real-time functionality check, a researcher used the system log data to document each time that a student should have been transitioned (anytime that they received an average self-explanation score below 2.0). This “researcher log” was then compared to the performance threshold feature logs. Overall, there was a 100% accuracy rate between the researcher and performance threshold feature logs (kappa of 1.0). This system check ensured that the performance threshold feature functioned correctly within the iSTART-2 system for this study.

Of the 28 students included in this study, 22 were transitioned at least once ($M = 1.64, SD = 1.50$). Thus, in at least one game, 22 of the 28 students had an average self-explanation score below 2.0 ($M = 0.93, SD = 0.70$). Although students were transitioned an average of 1.6 times, it is important to note that they were not forced to complete the remedial assignment (i.e., Coached Practice). Instead, students could choose to close out of this window and return to the game-based practice menu. In total, the 22 students were transitioned to the remedial Coached Practice 35 times, however, 76% of the time, the students chose to close out of the practice early and return to the game-based menu.

In-System Performance

A repeated- measures ANOVA was conducted to examine the effects of the performance threshold feature on self-explanation quality. This analysis examined how students’ self-

explanation scores changed between the game in which they scored below a 2.0 and the game directly after the transition. This analysis did not take into account whether students completed the remedial Coached Practice activity or their performance within Coached Practice, instead focusing on the self-explanation scores in the game *following* the remedial activity.

Of the 22 who were transitioned at least once, 16 completed a game following the transition and were included in this analysis. Students' average self-explanation scores prior to being transitioned were 0.96 ($SD = 0.70$). Conversely, students' average scores in the game following remedial Coached Practice was 1.99 ($SD = 0.69$), demonstrating a significant increase in self explanation scores ($F(1, 15) = 17.15, p < .001$). Although this analysis has a small sample size ($n = 16$), it achieves a high effect size (partial $\eta^2 = .533$) and a high-observed power of 0.97.

A follow-up repeated measures ANOVA was conducted using self-explanation quality for the games (prior to the transition and directly after the transition) as the within- subjects variable and students' choice to close out of the remedial function as a between- subjects variable (close out $n = 8$; complete $n = 8$). This analysis was designed to examine how students' choice to engage with the remedial function influenced any performance gains in self-explanation quality. Students who closed out of the remedial function on averaged scored a 1.96 ($SD = .72$) on their subsequent self-explanation, whereas students who completed the remedial coached practice scored an average 2.01 ($SD = .67$) on their subsequent self-explanation. This analysis revealed no significant effect of students' choice to close out on the average increase in self-explanation quality directly following the transition ($F(1,14) = .746, p = .402, \text{partial } \eta^2 = .051$). Thus, the

change in students' self-explanation scores directly after the transition did not depend on the completion of the remedial activity.

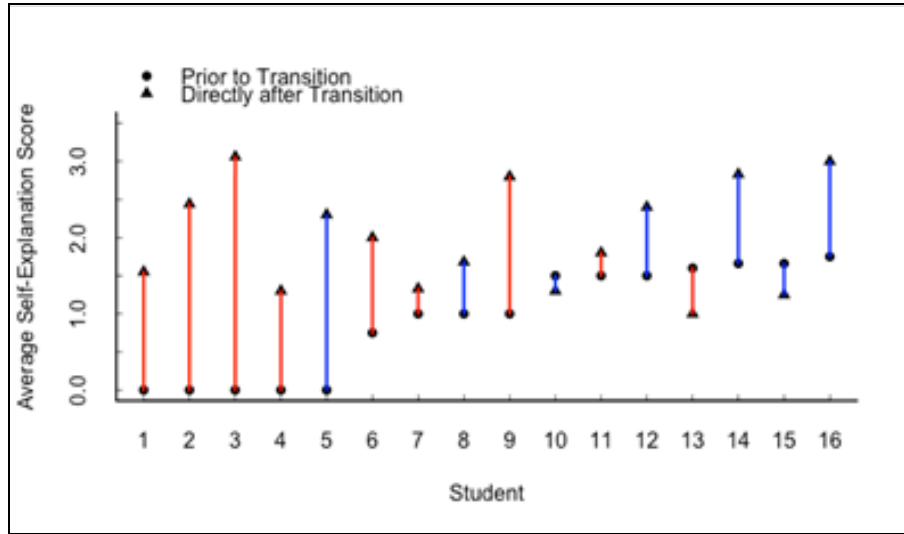


Figure 5. Transition gain scores per student. Red lines represent low prior knowledge students and blue lines represent high prior knowledge students.

Prior Knowledge. Finally, we examined how students' prior science knowledge related to transition gain scores. We define transition gain scores as the average self-explanation score a student received in games immediately following transitions minus the average self-explanation game score received immediately prior to the transitions. Using a correlation analysis, we found no significant relation between students' transition gain scores and their prior knowledge ($r = -.23, p = .386$). Thus, gains in students' self-explanation quality were not related to their prior domain knowledge in science. Figure 5 provides a visual representation of these gains as a function of prior knowledge. Red lines represent the low prior knowledge students ($M_{PK\ score} = 39\%, SD = 13\%$), and the blue lines represent the high prior knowledge students ($M_{PK\ score} = 78\%, SD = 11\%$), based on

a median split. Of the 16 students included in this analysis, 13 students (7 high knowledge and 9 low knowledge) showed improvements in the quality of their self-explanation scores after the transition.

STUDY 1 DISCUSSION

Study 1 examined how the transition function influenced strategy performance. Although Study 1 revealed that students showed gains in their self-explanation scores immediately following the remedial activity, students often chose not to complete that remedial activity and instead return to the game-based menu. Thus, participants' increase in self-explanation quality may be related to metacognition promoted by the pop-up message and not the direct instruction that they received in the remedial Coached Practice environment. Interestingly, the gains in self-explanation competency did not vary as a function of prior knowledge. That is, both high and low knowledge students benefited from the performance threshold feature. While Study 1 results were overwhelming positive, this work was still conducted on a very small sample size and thus in Study 2, we attempted to replicate the results found in Study 1. In Study 2, we also examined how removing students' option to close out of the remedial activity impacted learning gains. Thus, when a student was transitioned to Coached Practice because of poor performance, they were forced to complete remedial practice before being allowed to go back to the game-based menu. This manipulation affords us the opportunity to examine whether forcing students to complete the remedial activity (and thus taking away some level of agency) will produce similar learning gains to those found in Study 1.

STUDY 2

METHOD

Participants

Study 2 included 28 college students from a large university campus in the Southwest United States. These students were, on average, 19.4 years of age (range 18 to 25). Of the 28 students, 49% were male, 39% were Caucasian, 39% were Asian, 2% were African-American, 10% were Hispanic, and 10% reported other ethnicities.

Procedure

Similar to Study 1, Study 2 included a single 3-hour session consisting of a pretest, strategy training, game-based practice within iSTART-2, and a posttest. The procedures for Study 2 were similar to the Study 1 procedures. There were two important differences during the game-based practice. First students were able to self-select to interact with Coached Practice (which was not available in Study 1) in addition to Map Conquest and Showdown. Thus, students could choose to interact with any of these three practice environments at any rate or in any order that they chose. Importantly, students were only transitioned into the remedial Coached Practice environment if their average self-explanation score was below a 2.0 in either Map Conquest or Showdown; this is because students are prompted to self-explain multiple times in Coached Practice when their self-explanation is not of sufficient quality, making an average score less meaningful. The second difference was that students could no longer close out of the remedial Coached Practice activity as they could in Study 1.

Measures

The measures for Study 2 were identical to those used in Study 1. The one exception is that in Study 2 we added posttest system perception questions designed to gauge students' affect towards the transition message that students receive before they are sent to remedial Coached Practice (see Table 1).

Table 1.

Posttest Performance Threshold Feature Perception Questions

Transition Message Perceptions
The system pop-ups were annoying
The system pop-ups were frustrating
The system pop-ups were motivating
The system pop-ups were helpful
The system pop-ups helped me evaluate my performance
After receiving a system pop-up I tried harder
The system pop-ups were discouraging
After receiving a system pop-up I paid more attention to my performance

All Questions ranged from 1 (Strongly Disagree) to 6 (Strongly Agree)

STUDY 2 RESULTS

Games Played

Within the current study, students were free to interact with the game-based interface for approximately 1 hour. During this time, they could choose to play Coached Practice, Map Conquest, or Showdown as many times as they wanted and in whatever order they wished. On average, students completed 6.35 total game plays (range = 2 to 12, $SD = 2.54$), with Coached Practice being played an average of 3.14 times (range 1 to 7, $SD = 1.60$), Map Conquest being played an average of 1.82 times (range = 0 to 7, $SD = 1.66$),

and Showdown being played an average of 1.40 times (range 0 to 6, $SD = 1.34$). Bonferroni-adjusted pairwise comparisons indicated that Coached Practice was chosen at a significantly higher rate than Map Conquest ($t = 2.71, p^{\text{adjusted}} < .05$) and Showdown ($t = 4.56, p^{\text{adjusted}} < .01$). In contrast, an adjusted pairwise comparison indicated that there was no significant difference between how often Showdown was chosen compared to Map Conquest ($t = -1.07, p^{\text{adjusted}} > .05$).

Transitions

The current study builds off of Study 1 by examining the impact that the transition function had on students' in-game self-explanation quality. However, unlike Study 1, in Study 2, students had to complete remedial Coached Practice when they did not meet the 2.0 average score threshold. Of the 28 college students included within this study, 13 were transitioned at least once ($M = .93, SD = 1.09$). Thus, in at least one game, 13 students had an average self-explanation below 2.0 ($M = 1.17, SD = .50$).

In-System Performance

A repeated measures ANOVA was conducted to examine the effects of the performance threshold feature on self-explanation quality. This analysis was designed to examine how students' self-explanation scores changed in the game play directly after they were transitioned for the first time.

Of the 13 students who were transitioned at least once, 12 completed a game directly after being transitioned and therefore these 12 were included in this analysis. On average, students' self-explanation scores prior to being transitioned were 1.40 ($SD = .82$). Conversely, in the game after students completed the remedial Coached practice,

their average self-explanation score was a 2.19 ($SD = .67$). This is a significant increase in self explanation scores ($F(1, 11) = 14.60, p = .003$). Figure 6 provides a visual representation of these gains. Of the 12 students included in this analysis, 11 students showed improvements in the quality of their self-explanation scores after the transition. While again we have a small sample size ($n = 12$), we do achieve a high effect size (partial $\eta^2 = .570$) and a high-observed power of .94.

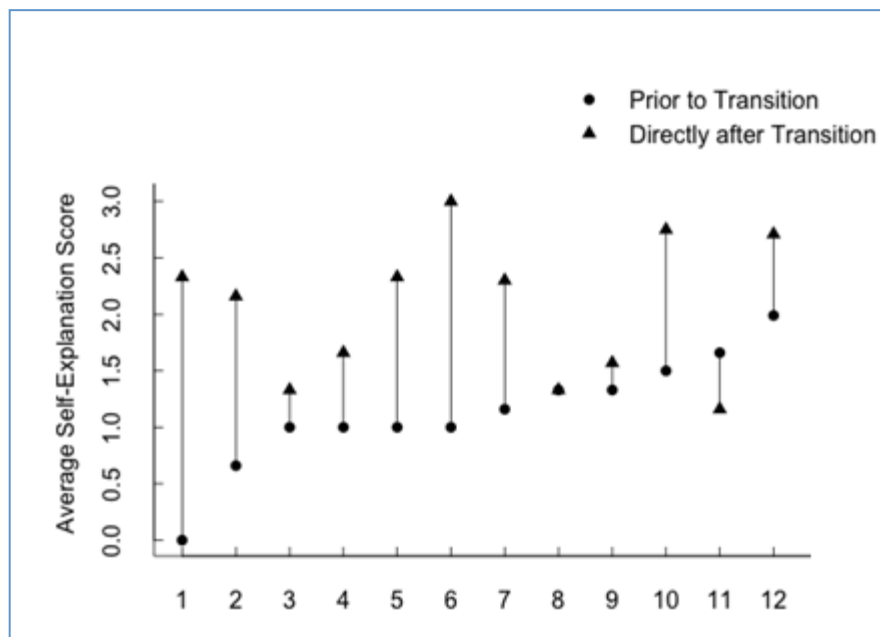


Figure 6. Transition gain scores per student.

Prior Knowledge. Finally, we were interested in examining how students' prior science knowledge related to transition gain scores. Transition gain scores are again defined as the average self-explanation score a student receives in the game immediately following the transition minus the self-explanation game score the students receives immediately prior to the transition. Using a correlation analysis, we found no significant relation between students' transition gain scores and their prior knowledge ($r = .496, p =$

.101). Thus, similar to Study 1, gains in students' self-explanation quality did not seem to be related to their prior science knowledge.

Posttest Performance Threshold Feature Perceptions

During posttest, students were asked to rate their perceptions of the transition messages that they encountered. Table 2 displays the descriptive statistics for each of the eight system perception questions. Overall, students seemed to rate the pop-up messages positively. Indeed, the three questions that were indicative of negative affect (popups were annoying, pop-ups were frustrating, and pop-us were discouraging) received the lowest scores. This suggests that the students viewed the transition message and feature favorably.

Table 2.

Posttest Performance Threshold Feature Perceptions

Transition Message Perceptions	Mean	SD	Range
The system pop-ups were annoying	3.07	0.94	2.0 - 5.0
The system pop-ups were frustrating	3.07	1.02	1.0 - 5.0
The system pop-ups were motivating	3.46	1.07	1.0 - 5.0
The system pop-ups were helpful	3.54	1.07	1.0 - 5.0
The system pop-ups helped me evaluate my performance	3.89	1.07	2.0 - 5.0
After receiving a system pop-up I tried harder	3.50	1.20	1.0 - 5.0
The system pop-ups were discouraging	2.64	0.87	1.0 - 4.0
After receiving a system pop-up I paid more attention to my performance	3.89	1.17	1.0 - 5.0

All Questions ranged from 1 (Strongly Disagree) to 6 (Strongly Agree)

STUDY 2 DISCUSSION

Study 2 replicated the results of Study 1 and revealed that students' self-explanation scores significantly improved after they received the transition message. Interestingly, it did not seem to matter whether or not students completed the remedial coached practice as both Study 1 and 2 saw significant changes in self-explanation quality. Results from Study 2 also replicated the results of Study 1 and revealed that both high and low knowledge students benefited from this feature. However, an additional question that needs to be addressed is whether changes in the threshold level may interact with students' prior knowledge to improve self-explanation quality. Thus, in Study 3 we manipulated the threshold level in order to examine if variations in the experimenter

threshold interacted with students' prior knowledge to impact in-system performance.

STUDY 3

METHOD

Participants

Study 3 included 84 college students from a large university campus in the Southwest United States. These students were, on average, 20.3 years of age (range 18 to 40). Of the 84 students, 60% were male, 47% were Caucasian, 27% were Asian, 8% were African-American, 10% were Hispanic, and 8% reported other ethnicities.

Procedure

Similar to Studies 1 and 2, Study 3 included a single 3-hour session consisting of a pretest, strategy training, game-based practice within iSTART-2, and a posttest. The procedures for Study 3 were identical to those of Study 2 for pretest, strategy training, and posttest. The only difference between Study 2 and 3 is that Study 3 included three threshold conditions. In Studies 1 and 2, students were transitioned to the remedial Coached Practice when their average self-explanation score did not reach the 2.0 threshold in Map Conquest or Showdown. In Study 3, students were randomly assigned to a 2.0, 2.5, or 2.75 threshold condition.

Measures

The measures used in Study 3 are identical to those used in Study 2.

STUDY 3 RESULTS

Games Played

In Study 3, students were free to interact with the game-based interface for approximately 1 hour. During this time, they could freely choose between Coached Practice, Map Conquest, or Showdown. On average, students completed 5.26 total game plays (range = 2 to 18, $SD = 3.07$), with Coached Practice played an average of 1.89 times (range 1 to 10, $SD = 1.55$), Map Conquest played average of 1.92 times (range = 0 to 10, $SD = 1.85$), and Showdown played an average of 1.44 times (range 0 to 5, $SD = 1.25$). Bonferroni-adjusted pairwise comparisons indicated that Coached Practice was chosen at a significantly higher rate than Showdown ($t = 2.24, p^{\text{adjusted}} < .05$) but not Map Conquest ($t = -.088, p^{\text{adjusted}} = .93$). An additional adjusted pairwise comparison indicated that Map Conquest was chosen significantly more often than Showdown ($t = 2.46, p^{\text{adjusted}} > .05$).

Transitions

Study 3 manipulated the threshold across three transition threshold conditions. These conditions included a low threshold of 2.0 ($n = 28$), a medium threshold of 2.5 ($n = 29$) and a high threshold of 2.75 ($n = 27$). Across all three conditions, 61 of the 84 participants were transitioned at least once ($M = 1.48, SD = 1.47$). In the low threshold condition, 20 of the 28 students were transitioned at least once ($M = 1.36, SD = 1.62$). In the medium threshold condition 21 of the 29 students were transitioned at least once ($M = 1.69, SD = 1.49$). Finally, in the high threshold condition 20 of the 27 students were transitioned at least once ($M = 1.49, SD = 1.47$). An ANOVA revealed no significant differences in times transitioned between the three conditions ($F(2, 83) = .415, p = .662$). One reason for this could be that when students were transitioned, on average, they were scoring very poorly (i.e., under a 2.0 for all conditions). Thus, the transition seems to most frequently impact low performing students.

In-System Performance

Overall. A repeated measures ANOVA was conducted to examine the effects of the transition function on self-explanation quality. This analysis is designed to examine how students' self-explanation scores changed in the game play after they were transitioned. Of the 61 students who were transitioned at least once, 52 completed a game (i.e., which provides performance data) immediately after being transitioned to the remedial Coached Practice and therefore were included in this analysis. On average, students' self-explanation scores prior to being transitioned were .79 ($SD = .68$). Conversely, after students completed the remedial Coached Practice, their average score on the next game play was a 1.93 ($SD = .95$). This is a significant increase in self explanation scores ($F(1, 51) = 66.32, p < .001$). Similar to the results from Studies 1 and 2, this effect achieved a high effect size (partial $\eta^2 = .565$) and a high observed power of 1.0.

By Condition. A second repeated measures ANOVA was conducted to examine the effects of the transition function on self-explanation quality as a function of threshold condition. This affords us the opportunity to examine how performance gains varied across conditions. This analysis revealed that there was no significant interaction between time (pre-transition, post transition) and threshold (high, medium, low) ($F(2, 49) = .186, p = .831$). Thus, the improvement in students' average self-explanation did not vary as a function of the transition threshold.

Prior Knowledge. Similar to Study 1 and 2, we were interested in examining how students' prior science knowledge related to transition gain scores. Transition gain

scores are again defined as the average self-explanation score a student receives in the game immediately following the transition minus the self-explanation game score the students receives immediately prior to the transition. Using a correlation analysis, we found no significant relation between students' transition gain scores and their prior knowledge ($r = .316, p = .271$). Thus, similar to Study 1 and 2, gains in students' self-explanation quality did not seem to be related to their prior science knowledge.

Finally, we examined how the transition threshold conditions interacted with students' prior science knowledge to impact changes in self-explanation quality. To conduct this analysis, we created a median split using students pretest prior knowledge scores, which resulted in two groups: high prior knowledge ($M = 7.50, SD = .75$) and low prior knowledge ($M = 4.65, SD = .71$). A repeated measures ANOVA was conducted to examine if the effects of the performance threshold feature on self-explanation quality were impacted by the interaction of prior knowledge and the transition threshold condition. This three-way interaction between time (pre-transition, post-transition), prior knowledge (high, low), and threshold setting (high, medium, low) was not significant ($F(2,46) = 2.29, p = .113$). Thus, students' self-explanation quality improved regardless of their prior knowledge or threshold condition.

Posttest Performance Threshold Feature Perceptions

Finally, we were interested in examining how the transition threshold conditions influenced students' posttest perceptions of the transition messages. Table 3 shows the means and standard deviations for all 8 questions across the 3 threshold conditions. Post-hoc Bonferroni-adjusted pairwise comparisons revealed that students in the high threshold condition viewed the messages as more helpful than students in the low

threshold condition ($p = .018$). Similarly, the high threshold students also reported that the messages aided them in evaluating their performance compared to students in the low threshold condition ($p = .018$). Finally, there was a marginal difference in students' reported effort after receiving a pop-up message. Specifically, students in the high threshold conditions reported that they tried harder after receiving the pop-up messages compared to the low threshold students ($p = .083$). These results suggest that students in the high threshold condition viewed the messages and performance threshold feature as more metacognitively helpful compared to the low threshold students. One reason for this may be that the high threshold could be seen as more challenging and thus push students to reflect more upon their performance.

Table 3.

Posttest Transition Message Perceptions per Condition

Transition Message Perceptions	Low Threshold	Medium Threshold	High Threshold
The system pop-ups were annoying	2.85 (1.37)	3.48 (1.35)	3.27 (.962)
The system pop-ups were frustrating	2.85 (1.22)	3.45 (1.24)	3.08 (1.07)
The system pop-ups were motivating	3.19 (1.67)	3.21 (1.21)	3.65 (.79)
The system pop-ups were helpful	3.27 (1.25)*	3.72 (.92)	3.85 (.73)*
The system pop-ups helped me evaluate my performance	3.27 (1.28)*	3.72 (1.22)	4.04 (.87)*
After receiving a system pop-up I tried harder	3.12 (1.31) <i>M</i>	3.55 (1.33)	3.69 (.84)
The system pop-ups were discouraging	2.73 (1.22)	3.00 (1.22)	2.96 (.82)
After receiving a system pop-up I paid more attention to my performance	3.38 (1.33)	3.52 (1.38)	3.73 (.87)

p<.05*, *p*<.10 *M*; **Mean (SD)**; All Questions ranged from 1 (Strongly Disagree) to 6 (Strongly Agree)

CHAPTER 2 DISCUSSION

Metacognitive skills are crucial for academic success. However, students vary in their ability to effectively control and regulate this behavior (Zimmerman, 1998). Recently, researchers have attempted to identify and develop more effective techniques designed to enhance students' metacognition during computer-based learning. The current paper is comprised of three studies that serve as functional and empirical tests of one such feature designed to indirectly promote metacognition during self-explanation practice. The overarching goal of this work is to provide a deeper understanding of how features embedded within adaptive environments can be leveraged to improve students' ability to reflect and monitor their own performance and subsequently improve learning outcomes.

This work had three primary research questions. The first of which, examined whether the performance threshold feature embedded within iSTART-2 was accurate in its use the online data from students' game plays to present messages and transition students to appropriate practice activities. The results presented here verify the functionality and accuracy of the performance threshold feature across all three studies. This new tool accurately collected and analyzed students' performance in real-time and subsequently transitioned students to a remedial practice when necessary. While this is by no means the end of the development of the performance threshold feature, the results presented here are promising for the further development of this functionality.

The second question examined how the performance threshold feature related to gains in students' self-explanation quality. Results also indicated that across all three studies, when students were presented with a pop-up message, their subsequent game

performance significantly improved. Moreover, students benefitted from this feature regardless of whether they completed the remedial activity. Thus, additional feedback was not the key to subsequent performance gains. These findings suggest that the performance threshold feature works as intended, promoting metacognition by informing and indirectly promoting reflection of students' performance and the goals of the system (or task). Thus, the performance threshold feature may be promoting metacognition by alerting students to their poor performance during the generative games. This assumption is further supported by the self-report surveys from Studies 2 and 3. Overall, students rated the messages favorably and stated that under higher performance expectations these messages promoted metacognitive processes. Specifically, results from Study 3 revealed that students in the high threshold condition viewed the transition messages as more helpful, aided them more in evaluating their performance, and motivated them to exert more effort compared to the students in the low threshold condition. This suggests that higher thresholds may promote more reflection and metacognitive processes. This may be due to the higher threshold being viewed as more challenging and thus require more reflection performance threshold feature compared to the low threshold students.

Finally, the third research question examined whether the impacts of the performance threshold feature varied as a function of individual differences in prior knowledge. Findings from the current studies revealed that when students were alerted to their poor performance and transitioned to the remedial practice environment, their subsequent performance was significantly better regardless of their prior knowledge. Thus, gains in learning were not solely influenced by the students' content knowledge. This is important as it suggest that students of varying skill levels can benefit from being

prompted to reflect upon the misalignment between their performance and the task standard (i.e., threshold). Combined, the results from all three studies reveal that the performance threshold feature within iSTART-2 may help to promote metacognition and positive learning outcomes.

CONCLUSION

The current work reveals promising insight into the development and testing of indirect metacognitive prompts embedded within adaptive environments. In the future, we will continue to test this new functionality in order to identify the optimal performance threshold and timing for delivering pop-up messages. Indeed, because students had been provided instruction on self-explanation strategies during the training phase of iSTART-2, a straightforward signal that they have room for improvement may have been enough to spur the use of those strategies. That is, the combination of direct and indirect support for metacognition within iSTART-2 seemed to be successful. Although students are expected to acquire reading strategies during the training phase (direct support), they did not consistently use them during practice, leading to low scores; however, the transition message (indirect support) caused immediate improvements in practice performance. We interpret this as evidence that the transition message encouraged monitoring behaviors, which subsequently led students to self-regulate and use strategies when self-explaining. This explanation of the benefits of indirect metacognitive support is consistent with previous work that positions indirect support as vital for learners who already have the necessary skills and strategies to succeed, but do not consistently use them (Bannert et al., 2009).

In the future, we also plan to examine the impact of this feature on

students' ability to exert agency over their learning path. The performance threshold feature temporarily removes students' ability to exert agency over their learning path. Our previous research, by contrast, has suggested that affording students choices promotes feelings of agency and ultimately improves learning outcomes (Snow, Allen, Jacovina, & McNamara, 2015; Snow, Jacovina, Allen, Dai, & McNamara, 2014). In the experiments presented here, students' learning paths were temporarily interrupted when they were transitioned to the remedial Coached Practice module. In the future, we will continue to investigate whether students should be afforded the choice to close out of the recommended module, or forced to complete it. Additional studies might examine the effects of specific transition message content (e.g., should the message be soft or harsh in its tone, or should it include strategy reminders), and how the messages may be adapted to students of particular skill levels or performance profiles.

The results presented here can expand beyond the current iSTART-2 system. Specifically, the results presented here have shown that alerting students to poor performance while also informing them of the standard that they should obtain can promote reflection and subsequent gains in learning. While this may seem like a simple feature embedded within a system, the performance threshold feature indirectly promoted metacognition and reflection. Thus, scientist interested in developing metacognitive support tools can use this work as a starting ground for their own development. The performance threshold feature within iSTART-2 is simple and yet very effective. In a world of constant change and technology, this work suggests that simple performance-based prompts can have huge impacts on a student's cognition and outcomes.

In conclusion, this study explored the use of a novel functionality within iSTART-2. The work presented here is our first attempt to build, test, and implement this feature in real-time. This methodology may be useful in the future for the improvement of student models that rely on understanding the relation between students' abilities and performance. Indeed, the tracking and modeling of performance trends and patterns over time is critical in our understanding of the ways that system features interact with students' metacognitive awareness, and ultimately their ability to learn from educational environments. Overall, these findings afford researchers and developers the opportunity to better understand how students' metacognitive awareness can be prompted and promoted within game-based learning environments.

CHAPTER 2 REFERENCES

- Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). Self-regulated learning with MetaTutor: Advancing the science of learning with MetaCognitive tools. *In New Science of Learning* (pp. 225-247). Springer New York.
- Azevedo, R., Witherspoon, A., Graesser, A.C., McNamara, D.S., Chauncey, A., Siler, E., Cai, Z., Rus, V., & Lintean, M. (2009). MetaTutor: Analyzing self-regulated learning in a tutoring system for biology. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A.C. Graesser (Eds.), *Artificial intelligence in education; Building learning systems that care; From knowledge representation to affective modeling* (pp. 635-637). Amsterdam, The Netherlands: IOS Press.
- Bannert, M., Hildebrand, M., & Mengelkamp, C. (2009). Effects of a metacognitive support device in learning environments. *Computers in Human Behavior*, 25, 829-835.
- Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, 31(2), 297-305.
- Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition.
- Brown, A. L., & DeLoache, J. S. (1978). Skills, plans, and self-regulation. *Children's thinking: What develops*, 3-35.
- Chi, M. T. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in instructional psychology*, 5, 161-238.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American psychologist*, 34(10), 906 -911.
- Graesser, A., & McNamara, D. S. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist*, 45(4), 234-244.
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (1998). *Metacognition in educational theory and practice*. Routledge.
- Jackson, G.T., & McNamara, D.S. (2011). Motivational impacts of a game-based intelligent tutoring system. In R. C. Murray & P. M. McCarthy (Eds.), *Proceedings of the 24th International Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 519-524). Menlo Park, CA: AAAI Press.
- Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based

- intelligent tutoring system. *Journal of Educational Psychology*, *105*, 1036-1049.
- Landauer, T. K., McNamara, D. S., Dennis, S. E., & Kintsch, W. E. (2007). Handbook of latent semantic analysis. Lawrence Erlbaum Associates Publishers.
- McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007). Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 227-241). Mahwah, NJ: Erlbaum.
- McNamara, D.S., & Kintsch, W. E. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247-288.
- Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal of Artificial Intelligence in Education*, *24*, 427-469.
- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology*, *25*, 92-104.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into practice*, *41*(4), 219-225.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational psychology review*, *7*(4), 351-371.
- Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading and Writing Quarterly*, *23*, 7-25.
- Snow, E. L., Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Does agency matter?: Exploring the impact of controlled behaviors within a game-based environment. *Computers & Education*, *26*, 378-392.
- Snow, E. L., Jackson, G. T., & McNamara, D. S. (2014). Emergent behaviors in computer-based learning environments: Computational signals of catching up. *Computers in Human Behavior*, *41*, 62-70.
- Snow, E. L., Jacovina, M. E., Allen, L. K., Dai, J., & McNamara, D. S. (2014). Entropy: A stealth assessment of agency in learning environments. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 241-244).
- Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system.

- Learning and Instruction*, 21, 267-280.
- Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Evaluative misalignment of 10th-grade student and teacher criteria for essay quality: An automated textual analysis. *Journal of Writing Research*, 5(2013), 35-59.
- Veenman, M. V., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and individual differences*, 15(2), 159-176.
- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *The Journal of Educational Research*, 84(1), 30-43.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166-183.
- Zimmerman, B. J., & Schunk, D. H. (2001). Reflections on theories of self-regulated learning and academic achievement. Self-regulated Learning and Academic Achievement: *Theoretical Perspectives*, 2, 289-307.

OVERALL DISCUSSION

Self-regulated learning (SRL) is related to students' ability to regulate, control, and monitor their behaviors while learning (Zimmerman, 2008) as well as to academic and professional success (Jarvela & Jarvenoja, 2011; Wolters, 2011). At the same time, research has shown that students vary in their ability to self-regulate (Zimmerman, 2008) and often need time to develop these skills (Zimmerman, 2008). The fluid nature of SRL has posed challenges for capturing changes in skill development across time (Greene & Azevedo, 2007). Traditionally, SRL has been measured through self-report measures that ask students to assess their own behaviors and intentions during learning tasks (Zimmerman, 2008). A concern that arises from the SRL literature is that these measures may not fully or adequately capture nuanced changes in behaviors that are often associated with the adaptive nature of SRL (Hadwin et al., 2007; McNamara, 2011; McNamara & Magliano, 2009). Indeed, it has been argued that there is often a *misalignment* between self-report measures and observable performance and behaviors (Hadwin et al., 2007; McNamara, 2011). This critique has led researchers to examine how users' behaviors and actions within adaptive environments may shed light upon the complex nature of the SRL process and its subcomponents (Hadwin et al., 2007; Snow et al., 2015). Relevant to the current work, trace data analyses have been shown to capture fine-grained variations in students' learning behaviors that would have otherwise been missed through the use of traditional self-reports (Hadwin et al., 2007; Snow et al., 2015). The work presented in this dissertation builds upon this premise by using both post-hoc (Chapter 1) and real-time (Chapter 2) trace data analyses to understand and promote sub-components of SRL within a game-based environment.

Chapter 1 investigated the relations between traditional measures of control and regulation (i.e., self-reports) and post-hoc trace data analyses. In particular, this work focused on visualizing and classifying *strategic patterns* that manifest in students' behaviors over time and then relating those patterns to one aspect of self-regulation, agency. The results from this work revealed that controlled patterns of behavior were significantly related to students' self-reported feelings of agency at posttest. That is, students who acted in a more decisive manner during training also reported feeling that they had more control over the system during training. It was also found that students who acted in a more controlled manner were more likely to perform better in the system than those who acted in a random manner. In follow-up Study 2, controlled and random patterns from Study 1 were yoked to students in Study 2. Interestingly, when this manipulation occurred the difference in performance scores between controlled and random patterns disappeared. Thus, these results do not seem to be driven simply by optimal trajectories or "gaming behavior" but instead, it seems that agency was a key component of success within the system and that trace data analyses could identify a data signature of agentic behavior (Snow et al., 2015).

Chapter 1 shed light upon the utility of log data and its ability to covertly assess behaviors associated with SRL. However, as stated earlier, the ultimate goal of this research is to use online measures in real time to guide students toward optimal learning behaviors. Chapter 2 examined one potential way to provide this real-time guidance through the combination of students' in-game performance data and an experimenter-set threshold. In a series of studies, it was found that students' in-system performance improved after they were alerted to their poor performance and then transitioned into a

remedial activity. Further, these results did not vary as a function of individual differences in prior knowledge. Thus, both high and low ability students' self-explanations improved when they were exposed to the performance threshold feature built into iSTART-2. Interestingly, students who interacted with the performance threshold feature reported that the feature helped them reflect and monitor their performance during game-play. Combined, these results tentatively suggest that the performance threshold feature may indirectly promote metacognition thus leading to enhanced in-system performance.

While the results presented in Chapter 2 are promising, two limitations are notable. First, the sample sizes for both Study 1 ($n=28$) and Study 2 ($n=28$) are relatively small and therefore comparisons between the two studies should be interpreted with caution. Although there seems to be no significant differences in mean gain scores comparing Study 1 and Study 2, with a larger sample size, the mean differences may prove to be more apparent. Nonetheless, if a comparison were to be made between the two studies, the sample size affords an observed power of .24 to detect a small effect size.

A second concern regards claims about the effects of the performance threshold feature on students' metacognitive awareness. Clearly, further empirical evaluation, specifically an experimental study that includes a control condition, is necessary prior to drawing strong conclusions. Such a study would afford the opportunity to examine how the performance threshold feature impacts self-reported metacognitive awareness compared to students who are not exposed to the feature.

Overall, the work presented in Chapter 2 is designed to contribute to our understanding of how embedded system features that are guided by real-time analytics

can be used to optimize pedagogy and learning outcomes within ITSs. Specifically, these features can be used to help students regulate and monitor their own performance. Within the SRL literature, it is stated that many students are unaware of how to set strategic plans during learning tasks (Zimmerman, 2008). Thus, system features may be one way to prompt students' metacognitive awareness. Previous work has shown that students are able to improve their SRL ability with instruction (Muraven, Baumeister, & Tice, 1999). Indeed, the use of real-time analytics to guide system features may help students become more aware of their behaviors and provide them with feedback on how to optimize their time within the system.

Both post-hoc (Chapter 1) and real-time (Chapter 2) trace data analyses are fundamental to our understanding of the SRL process. Post-hoc analyses serve as a way to *reverse engineer* the SRL process. That is, they can provide researchers or educators with information on *how* students regulate themselves. These measures are unobtrusive and provide a window into the learning process that would have otherwise been missed through more traditional (i.e., self-report) metrics. While post-hoc analyses often do not provide meaningful feedback during the learning task, they can be used to guide future students toward “optimal” learning trajectories or strategies within an adaptive environment. Indeed, these methodologies may prove useful for the improvement of student models that rely on understanding the relation between students' abilities and performance. The tracking and modeling of behavioral trends and patterns over time is critical to our understanding and ability to effectively support the various ways in which students self-regulate during learning.

It is important to note that the use of post-hoc analyses alone will not provide

researchers with a complete understanding of how to best support SRL and its subcomponents. A complimentary approach is to use real-time analytics to guide the design of system features that support various regulatory processes (Snow et al., under review). These techniques can be used to examine how information gleaned from trace data can help students improve their SRL skill. Indeed, real-time analytics can be used to help students better understand their SRL processes and provide feedback that they can use to develop their SRL abilities. To this end, real-time analytics can guide the design and visualization of embedded system tools that can pilot students toward optimal SRL behaviors.

Adaptive environments have the power to capture fluctuations in behaviors and actions through the use of trace data and both post-hoc and real-time analytics. These fine-grained data structures have the ability to shed light upon the complex nature of SRL. However, further work is needed to examine the extent to which various methodologies *best* capture nuanced changes in behaviors and actions across multiple systems. Currently, there are numerous metrics and methods used to capture fine-grained changes in behavior. One criticism, however, is that there are too many methodologies and a lack of systematic comparison between them. Along these lines, Greene and Azevedo, (2007) argued that there has been a lack of strict scientific examinations of the systems and methods used to capture and promote SRL. Thus, system designers often implement a variety of features and methodologies without explicitly testing their effects on learning and SRL processes. While learning analytics and its associated techniques can be seen as exciting and promising, many of these methodologies are still in their infancy and therefore more theoretically guided work is needed to improve the

development and testing of such methodologies and system features. One way, to approach this issue is through the identification of generalizable methodologies that can be employed to measure SRL across a variety of systems and data sets. The development of generalizable methodologies will aid in the development of construct validity and reliability for the analyses and implementation of both post-hoc and real-time analytics designed to capture and promote SRL and its sub-components.

In conclusion, the work presented in this dissertation is among the first to use post-hoc trace-data analyses to investigate *how* users approach various learning tasks and real-time performance analyses to guide system pedagogy. The goal of this work is to identify and promote SRL behaviors within adaptive environments. The results and methodologies presented here serve as a starting point for scientists interested in analyzing users' interactions within game-based environments in order to gain a deeper understanding of the various aspects of students' regulatory processes.

COMPREHENSIVE REFERENCES

- Allen, L. K., Snow, E.L., & McNamara, D. S. (2014). The long and winding road: Investigating the differential writing patterns of high and low skilled writers. In J. Stamper, S. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 304-307). London, UK.
- Azevedo, R., Johnson, A., Chauncey, A., & Burkett, C. (2010). Self-regulated learning with MetaTutor: Advancing the science of learning with MetaCognitive tools. *In New Science of Learning* (pp. 225-247). Springer New York.
- Azevedo, R., Witherspoon, A., Graesser, A.C., McNamara, D.S., Chauncey, A., Siler, E., Cai, Z., Rus, V., & Lintean, M. (2009). MetaTutor: Analyzing self-regulated learning in a tutoring system for biology. In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A.C. Graesser (Eds.), *Artificial intelligence in education; Building learning systems that care; From knowledge representation to affective modeling* (pp. 635-637). Amsterdam, The Netherlands: IOS Press.
- Bandura, A. (1989). Human agency in social cognitive theory. *American psychologist*, 44(9), 1175 – 1184.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual review of psychology*, 52(1), 1-26.
- Bannert, M., Hildebrand, M., & Mengelkamp, C. (2009). Effects of a metacognitive support device in learning environments. *Computers in Human Behavior*, 25, 829-835.
- Barab, S., Pettyjohn, P., Gresalfi, M., Volk, C., & Solomou, M. (2012). Game-based curriculum and transformational play: Designing to meaningfully positioning person, content, and context. *Computers & Education*, 58(1), 518-533.
- Benhamou, S., & Bovet, P. (1989). How animals use their environment: A new look at kinesis. *Animal Behavior*, 38, 375-383.
- Benjamin, A. S. (2003). Predicting and postdicting the effects of word frequency on memory. *Memory & Cognition*, 31(2), 297-305.
- Bishop, G. F. (1987). Experiments with the middle response alternative in survey questions. *Public Opinion Quarterly*, 51(2), 220-232.
- Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition.

- Brown, A. L., & DeLoache, J. S. (1978). Skills, plans, and self-regulation. *Children's thinking: What develops*, 3-35.
- Calvert, S., Strong, B., & Gallagher, L. (2005). Control as an engagement feature of young children's attention to and learning of computer content. *The American Behavioral Scientist*, 48, 578-589.
- Chi, M. T. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. *Advances in instructional psychology*, 5, 161-238.
- Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2009). Combining shared control with variability over surface features: Effects on transfer test performance and task involvement. *Computers in Human Behavior*, 25, 290–298.
- Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2011). Learner-controlled selection of tasks with different surface and structural features: Effects on transfer and efficiency. *Computers in Human Behavior*, 27, 76–81.
- Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88, 715– 730.
- Davey, H. M., Barratt, A. L., Butow, P. N., & Deeks, J. J. (2007). A one-item question with a Likert or Visual Analog Scale adequately measured current anxiety. *Journal of clinical epidemiology*, 60, 356-360.
- De Boer, A. G. E. M., Van Lanschot, J. J. B., Stalmeier, P. F. M., Van Sandick, J. W., Hulscher, J. B. F., De Haes, J. C. J. M., & Sprangers, M. A. G. (2004). Is a single-item visual analogue scale as valid, reliable and responsive as multi-item scales in measuring quality of life?. *Quality of Life Research*, 13(2), 311-320.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press, New York.
- Ellis, D., & Zimmerman, B. J. (2001). Enhancing self-monitoring during self-regulated learning of speech. In H. J. Hartman (Ed.), *Metacognition in learning and instruction. Theory, research, and practice* (pp. 205–228). Dordrecht: Kluwer Academic Press.
- Fasolo, B., Hertwig, R., Huber, M., & Ludwig, M. (2009). Size, entropy, and density: What is the difference that makes the difference between small and large real-world assortments? *Psychology & Marketing*, 26, 254-279.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–

- developmental inquiry. *American psychologist*, 34(10), 906 -911.
- Flowerday, T., & Schraw, G. (2000). Teachers' beliefs about instructional choice: A phenomenological study. *Journal of Educational Psychology*, 92, 634–645.
- Flowerday, T., & Schraw, G. (2003). Effect of choice on cognitive and affective engagement. *Journal of Educational Research*, 96, 207–215.
- Flowerday, T., Schraw, G., & Stevens, J. (2004). The role of choice and interest in reader engagement. *Journal of Experimental Education*, 72, 93–114.
- Ford, M. E. (1992). *Motivating humans: Goals, emotions, and personal agency beliefs*. Newbury Park, CA: Sage Publications.
- Graesser, A., & McNamara, D. S. (2010). Self-regulated learning in learning environments with pedagogical agents that interact in natural language. *Educational Psychologist*, 45(4), 234-244.
- Greene, J. A., & Azevedo, R. (2007). A Theoretical Review of Winne and Hadwin's Model of Self-Regulated Learning: New Perspectives and Directions. *Review of Educational Research*, 334-372.
- Grossman, E. R. F. W. (1953). Entropy and choice time: The effect of frequency unbalance on choice-response. *Quarterly Journal of Experimental Psychology*, 5(2), 41-51.
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (1998). *Metacognition in educational theory and practice*. Routledge.
- Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, 2, 107-124.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41, 111-127.
- Iyengar, S., & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79, 995–1006.
- Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology*, 105, 1036-1049.
- Jackson, G. T., Guess, R. H., & McNamara, D. S. (2009). Assessing cognitively complex strategy use in an untrained domain. *Topics in Cognitive Science*, 2, 127-137.
- Jackson, G.T., & McNamara, D.S. (2011). Motivational impacts of a game-based

- intelligent tutoring system. In R. C. Murray & P. M. McCarthy (Eds.), *Proceedings of the 24th International Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 519-524). Menlo Park, CA: AAAI Press.
- Järvelä, S., & Järvenoja, H. (2011). Socially constructed self-regulated learning and motivation regulation in collaborative learning groups. *Teachers College Record, 113*, 350-374.
- Katz, I., Assor, A., Kanat-Maymon, Y., & Bereby-Meyer, Y. (2006). Interest as a motivational resource: Feedback and gender matter, but interest makes the difference. *Social Psychology of Education, 9*, 27– 42.
- Klopper, F. J., & Madden, T. M. (1980). The Middlemost Choice on Attitude Items: Ambivalence, Neutrality, or Uncertainty?. *Personality and Social Psychology Bulletin, 6*(1), 97-101.
- Landauer, T. K., McNamara, D. S., Dennis, S. E., & Kintsch, W. E. (2007). *Handbook of latent semantic analysis*. Lawrence Erlbaum Associates Publishers.
- Lester, J. C., Mott, B. W., Robison, J. L., Rowe, J. P., & Shores, L. R. (2013). Supporting self-regulated science learning in narrative-centered learning environments. *In International handbook of metacognition and learning technologies* (pp. 471-483). Springer New York.
- Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biological Evolution, 13*, 660–665.
- McIver, J. P., & Carmines, E. G. (1981). *Unidimensional scaling*. Thousand Oaks, CA: Sage.
- McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: Challenges and successes. *Megacognition and Learning, 3*, 1-11.
- McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007). Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 227-241). Mahwah, NJ: Erlbaum.
- McNamara, D. S., Jackson, G. T., & Graesser, A. C. (2010). Intelligent tutoring and games (ITaG). In Y. K. Baek (Ed.), *Gaming for classroom-based learning: Digital role-playing as a motivator of study* (pp. 44-65). Hershey, PA: IGI Global.
- McNamara, D.S., & Kintsch, W. E. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes, 22*, 247-288.

- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers*, 36, 222-233.
- McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I. B. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 397-420). Mahwah, NJ: Erlbaum.
- McNamara, D.S., & Magliano, J. P. (2009). Self-explanation and metacognition: The dynamics of reading. In J.D. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 60-81). Mahwah, NJ: Erlbaum.
- McNamara, D. S., & Shapiro, A. M. (2005). Multimedia and hypermedia solutions for promoting metacognitive engagement, coherence, and learning. *Journal of Educational Computing Research*, 33, 1-29.
- Metcalfe, J., Eich, T. S., & Miele, D. B. (2013). Metacognition of agency: Proximal action and distal outcome. *Experimental Brain Research*, 229, 485–96.
- Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & Quantity*, 42(6), 779-794.
- Muraven, M., Baumeister, R. F., & Tice, D. M. (1999). Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise. *The Journal of Social Psychology*, 139(4), 446-457.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3) McGraw-Hill. New York.
- Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal of Artificial Intelligence in Education*, 24, 427-469.
- O'Reilly, T.P., Sinclair, G.P., & McNamara, D.S. (2004). Reading strategy training: Automated versus live. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1059-1064). Mahwah, NJ: Erlbaum.
- Parker, L. E., & Lepper, M. R. (1992). Effects of fantasy contexts on children's learning and motivation: Making learning more fun. *Journal of Personality and Social Psychology*, 62, 625–633.
- Paulhus, D. L., Lysy, D. C., & Yik, M. S. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests?. *Journal of Personality*, 66, 525-554.

- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychological Review, 18*, 315-341.
- 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*, 531-549.
- Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25*, 92-104.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into practice, 41*(4), 219-225.
- Reynolds, P. L., & Symons, S. (2001). Motivational variables and children's text search. *Journal of Educational Psychology, 93*, 14-22.
- Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring global self-esteem: Construct validation of a single-item measure and the Rosenberg Self-Esteem Scale. *Personality and social psychology bulletin, 27*, 151-161.
- Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*, 267-280.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review, 16* (2), 225-237.
- Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2010). Individual differences in gameplay and learning: A narrative-centered learning perspective. *In Proceedings of the Fifth International Conference on the Foundations of Digital Games* (pp. 171-178). ACM.
- Sabourin, J., Shores, L. R., Mott, B. W., & Lester, J. C. (2012). Predicting student self-regulation strategies in game-based learning environments. *In Intelligent Tutoring Systems* (pp. 141-150). Springer Berlin Heidelberg.
- Schmierbach, M., Limperos, A. M., & Woolley, J. K. (2012). Feeling the need for (personalized) speed: How natural controls and customization contribute to enjoyment of a racing game through enhanced immersion. *Cyberpsychology, Behavior, and Social Networking, 15*(7), 364-369.

- Schönau-Fog, H., & Bjørner, T. (2012). "Sure, I would like to continue": A method for mapping the experience of engagement in video games. *Bulletin of Science, Technology & Society*, 32, 405–412.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational psychology review*, 7(4), 351-371.
- Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading and Writing Quarterly*, 23, 7-25.
- Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, 30, 50–64.
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer Games and Instruction*, 55, 503-524.
- Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. *In Serious games: Mechanisms and effects*. (pp. 295-321). Philadelphia, PA: Routledge/LEA, 2.
- Spires, H. A., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). Problem solving and game-based learning: Effects of middle grade students' hypothesis testing strategies on learning outcomes. *Journal of Educational Computing Research*, 44(4), 453-472.
- Snow, E. L., Allen, L. K., Jackson, G. T., & McNamara, D. S. (2014). Tracking choices: Computational analysis of learning trajectories. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 316-319). Heidelberg, Berlin, Germany: Springer.
- Snow, E. L., Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Does agency matter?: Exploring the impact of controlled behaviors within a game-based environment. *Computers & Education*, 26, 378-392.
- Snow, E. L., Allen L. K., Russell, D. G., & McNamara, D. S. (2014). Who's in control?: Categorizing nuanced patterns of behaviors within a game-based intelligent tutoring system. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 185-192). Heidelberg, Berlin, Germany: Springer.
- Snow, E. L., Jackson, G. T., & McNamara, D. S. (2014). Emergent behaviors in computer-based learning environments: Computational signals of catching up. *Computers in Human Behavior*, 41, 62-70.

- Snow, E. L., Jacovina, M. E., Allen, L. K., Dai, J., & McNamara, D. S. (2014). Entropy: A stealth assessment of agency in learning environments. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 241-244). Heidelberg, Berlin, Germany: Springer.
- Snow, E. L., Jacovina, M. E., Allen, L. K., & McNamara, D. S. (under review). Promoting metacognitive awareness within a game-based environment. *Manuscript submitted to Computers and Education*.
- Snow, E. L., Jackson, G. T., Varner, L. K., & McNamara, D. S. (2013). Investigating the effects of off-task personalization on system performance and attitudes within a game-based environment. In S. K. D'Mello, R. A. Calvo, and A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 272-275). Heidelberg, Berlin, Germany: Springer.
- Snow, E. L., Likens, A., Jackson, G. T., & McNamara, D. S. (2013). Students' walk through tutoring: Using a random walk analysis to profile students. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 276-279). Heidelberg, Berlin, Germany: Springer.
- Stone, M. H. (2004). Substantive scale construction. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp.201–225). Maple Grove, MN: JAM Press.
- Tabbers, H. K., & de Koeijer, B. (2010). Learner control in animated multimedia instructions. *Instructional Science*, 38, 441–453.
- Taylor, R. S., O'Reilly, T., Rowe, M., & McNamara, D. S. (2006). Improving understanding of science texts: iSTART strategy training vs. web design control task. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 2234-2239). Mahwah, NJ: Erlbaum.
- Teng, C. I. (2010). Customization, immersion satisfaction, and online gamer loyalty. *Computers in Human Behavior*, 26, 1547–1554.
- Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Evaluative misalignment of 10th-grade student and teacher criteria for essay quality: An automated textual analysis. *Journal of Writing Research*, 5(2013), 35-59.
- Veenman, M. V., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and individual differences*, 15(2), 159-176.

- Wang, M. C., Haertel, G. D., & Walberg, H. J. (1990). What influences learning? A content analysis of review literature. *The Journal of Educational Research*, 84(1), 30-43.
- Wolters, C.A. (2011). Regulation of motivation: Contextual and social aspects. *Teachers College Record*, 113, 265-283.
- Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, 9(6), 772-775.
- Zhou, M. (2013). Using traces to investigate self-regulatory activities: A study of self-regulation and achievement goal profiles in the context of web search for academic tasks. *Journal of Cognitive Education and Psychology*, 12, 287-305.
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25, 3-17.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166-183.
- Zimmerman, B. J., & Schunk, D. H. (2001). Reflections on theories of self-regulated learning and academic achievement. *Self-regulated Learning and Academic Achievement: Theoretical Perspectives*, 2, 289-307.