

Stealth Assessment of Self-Regulative Behaviors within a Game-Based Environment

by

Erica L. Snow

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved May 2014 by the
Graduate Supervisory Committee:

Danielle McNamara PhD, Chair
Arthur Glenberg PhD
Nick Duran PhD

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

Students' ability to regulate and control their behaviors during learning has been shown to be a critical skill for academic success. However, researchers often struggle with ways to capture the nuances of this ability, often solely relying on self-report measures. This thesis proposal employs a novel approach to investigating variations in students' ability to self-regulate by using process data from the game-based Intelligent Tutoring System (ITS) iSTART-ME. This approach affords a nuanced examination of how students' regulate their interactions with game-based features at both a coarse-grained and fine-grain levels and the ultimate impact that those behaviors have on in-system performance and learning outcomes (i.e., self-explanation quality). This thesis is comprised of two submitted manuscripts that examined how a group of 40 high school students chose to engage with game-based features and how those interactions influenced their target skill performance. Findings suggest that in-system log data has the potential to provide stealth assessments of students' self-regulation while learning.

TABLE OF CONTENTS

	Page
LIST OF TABLES.....	V
LIST OF FIGURES.....	VI
CHAPTER	
1 INTRODUCTION.....	1
Intelligent Tutoring Systems.....	2
Thesis Project.....	5
2 STUDY 1.....	8
Introduction.....	8
iSTART.....	12
iSTART-ME.....	15
Current Study.....	18
Methods.....	19
Results.....	23
Discussion.....	29
Conclusion.....	33
3 STUDY 2.....	34
Introduction.....	34
iSTART-ME.....	39
Current Study.....	43

CHAPTER	Page
Method	44
Quantitative Methods	46
Results	51
Discussion.....	54
4 GENERAL DISCUSSION.....	58
Future Directions and Follow-Up Studies.....	63
Conclusion	77
REFERENCES	79

LIST OF TABLES

Table	Page
1. Cost of System Interaction in iSTART-ME – Study 1	17
2. Daily Enjoyment and Motivation Questions– Study 1	21
3. Game Achievement and Spendency Regression Outcomes– Study 1	26
4. Strategy Performance and Spendency Correlations– Study 1	27
5. Strategy Performance and Spendency Regression Outcomes– Study 1	28
6. Learning Transfer Regression Outcomes – Study 1	29
7. System Interaction Axis Assignment – Study 2	47
8. Target Skill Correlations– Study 2	52
9. Daily Strategy Regression Outcomes– Study 2.....	53
10. Long-Term Learning Regression Outcomes– Study 2.....	54
11. Pop-Up Messages – Follow-Up Study 3	74

LIST OF FIGURES

Figure	Page
1. Screen Shot of Demonstration Module - Study 1	14
2. Screen Shot of iSTART-ME Selection Menu - Study 1	15
3. Screen Shot of iSTART-ME Achievement Screen - Study 1.....	16
4. Avatar Configurations - Study 1	18
5. Balloon Bust - Study 1	18
6. Spendency Interaction Pattern - Study 1	25
7. Screen Shot of iSTART-ME Selection Menu - Study 2	41
8. Screen Shot of Map Conquest - Study 2	42
9. Balloon Bust - Study 2.....	43
10. Example Random Walk - Study 2.....	48
11. Student Generated Random Walk - Study 2	49
12. Student Generated Random Walk - Study 2	49
13. Transitional Probability of High Hurst Student	61
14. Transitional Probability of Low Hurst Student.....	62
15. Factorial Design Follow-Up Study3	72

Chapter 1

INTRODUCTION

Students' behaviors during a learning task can shed light upon the way in which they approach their learning process. For instance, some students may set goals and plans before starting a task, thus, revealing a decisive and strategic approach to their learning process (McCombs, 1989). Conversely, other students can approach the same task with no discernible scheme or method, thus, revealing an impetuous approach to their learning process. The emergence of students' ability to set decisive goals, plans, and make decisions during a task is often referred to as self-regulation (Zimmerman, 1990). Self-regulated learning (SRL) has been shown to have positive effects on learning and cognitive states (Jarvela & Jarvenoja, 2011; Winne & Hadwin, 2013; Wolters, 2011; Zimmerman, 2008). Collectively, SRL research has revealed that learning outcomes are enhanced when students more accurately monitor and self-assess their learning and performance (Ariel, Dunlosky, & Bailey, 2009; Harris et al., 2005).

One important component of self-regulation is students' ability to control their choices and set forth a strategic plan of action while engaging in a learning task (Schunk & Zimmerman, 2003). Strategic actions can range from various studying techniques (Hadwin & Winne, 2007) to game-play strategies (Sabourin et al., 2012). Zimmerman (1990) proposed that when students take personal responsibility over their learning experiences by creating strategic plans, they are more likely to succeed than those students who do not. Thus, students who do not exert control over their own learning behaviors often stifle their academic success (McCombs, 1989).

Importantly, self-regulated learning and its components (i.e., strategic planning) are not static (i.e., unchanging), instead researchers have shown this ability is dynamic in nature and evolves overtime (Boekaerts, Pintrich, & Zeidner, 2000; Hadwin et al., 2007; Zhou, 2013). This

constant fluctuation makes self-regulation and its corresponding components difficult to assess. Traditionally, self-regulation has been measured through the use of self-reports. One concern about using these self-report measures is that they may not fully or adequately capture the complex construct of SRL (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; Zhou, 2013). Thus, if self-regulation is indeed a dynamic and fluid skill that evolves over time, the use of such static measures may be hamstringing our understanding of students' learning process.

An alternative to the use of self-reports may be the utilization of online measures. In contrast to *offline* self-report and post assessments, online measures capture behaviors (e.g., language input and choices patterns) from the learner during their performance on various learning tasks (Hadwin et al., 2007; McNamara & Magliano, 2009; Winne & Hadwin, 2013). Online measures may be particularly informative of SRL because they can focus on fine-grained patterns in students' behaviors during learning and thus, may be more likely to capture how students exert agency while engaging in learning tasks.

Intelligent Tutoring Systems

Game-based Intelligent Tutoring Systems (ITS) potentially offer a venue for the use of online SRL measures. Intelligent Tutoring Systems (ITSs) are sophisticated learning environments that provide customized instruction to students based on their individual needs (Murray, 1999). These systems are more advanced than traditional computer-assisted training, in that they adapt to the users' performance and skill levels (Shute & Psotka, 1994). The adaptive nature of ITSs has resulted in the successful integration of these systems into a variety of settings (Bratt, 2009; Johnson & Valente, 2009; Lynch, Ashley, Alevan, & Pinkwart, 2006; Rai & Beck, 2012; Siemer & Angelides, 1994). One hypothesized explanation for the success of these systems is that ITSs provide customized feedback to students based on their performance in the system. This

feedback and customization allows students to progress through learning tasks at a pace that is appropriate to their customized learning model (Alevan & Koedinger, 2002). It also ensures that students are not only learning at a shallow procedural level, but they are gaining deeper knowledge at an appropriate pace.

Although ITSs effectively improve students' performance and promote overall learning, they have limitations. For instance, tutoring systems that specialize in complex skill and strategy acquisition may require extensive amounts of training and practice (McNamara, Jackson, & Graesser, 2009). These sessions can often require students to complete redundant tasks, which can lead to students' disengagement (Bell & McNamara, 2007; Jackson & McNamara, 2013). The increase of student disengagement over extended periods of time has led researchers to examine how the implementation of game-based features can enhance student motivation and engagement within learning systems (Cordova & Lepper, 1996; McNamara, et al., 2009; Rai & Beck, 2012).

Researchers have integrated game-based features into ITSs as a way to engage students and maintain their interest overtime. Rai, Beck, and Heffernan (2010) have defined game-based features as any element within the system that engages students' interest through incentives (e.g., points, badges), rewards (e.g., unlocking features based on performance or trophies), or interactivity (e.g., choice, customization). When game-based features are incorporated into an educational learning environment, students report an increase in motivation and engagement (Baker et al., 2008; Cordova & Lepper, 1996; Facer et al., 2004; Jackson & McNamara, 2011; Rai & Beck, 2012; Sabourin, Rowe, Mott, & Lester, 2011). For instance, Rai, Beck, and Heffernan found that when compared to a regular tutor, students preferred a tutor with game-based elements. Similarly, McQuiggan, Rowe, Lee, and Lester (2008) found that when game-

based features were added to an adaptive learning environment, students showed an increase in interest compared to traditional instruction.

Game-based environments provide researchers with an opportunity to incorporate *stealth* assessments of learning behaviors. These assessments are often embedded within the environment and are essentially “invisible” to students (Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Shute, 2011). One approach to stealth assessment is to embed assessments within the learning environment, and collect various performance measures while students interact with the various game-based features within the system interface. Another, and the one we evaluate in this study, is the use of log-data, also referred to as keystroke, mouse click, click stream, or telemetry data (depending on the context). Log-data is essentially the recording of all of a user’s interactions or keystrokes while interacting with various features in the system. Notably, the collection of log-data is not *built into* all computerized systems, but rather must be intentionally programmed. When it is collected, log-data can provide a wealth of information, particularly concerning students’ choices and *agency* while engaged with a system.

Log-data from game-based environments has been shown to be a useful tool in the detection of SRL behaviors. For instance, Sabourin and colleagues (2012) examined how log-data from the narrative-centered environment, Crystal Island, was indicative of students’ SRL strategy use (e.g., self-monitoring and goal setting). Sabourin et al. investigated how students’ behaviors during game-play (e.g., use of notes, books, or in-game tests) and pretest self-report measures of affect and prior knowledge combined to predict students’ level (i.e., low, medium, or high) of SRL strategy use. They found that the inclusion of system log-data within their models significantly contributed to the accurate classification of students’ use of SRL strategies. Such research demonstrates that log-data extracted from a game-based environments yield an

unobtrusive opportunity to examine variations in the way in which students regulate their behaviors during learning tasks.

Thesis Project

The proposed thesis project is comprised of two submitted journal manuscripts. Both of these manuscripts use system log-data, to investigate how students' regulate their behaviors within the game-based environment, iSTART-ME. Combined these two manuscripts examine students' interactions with game-based features at two levels: coarse-grained (total frequency) and fine-grained (nuanced pattern). A novel contribution of this thesis stems primarily from the analysis of process data to investigate how students regulate their interactions with different types of game-based features, and the subsequent impact these interactions have on learning outcomes. Thus, the goal of the combined work is to gain a deeper understanding of how students' regulate their behaviors within adaptive systems and how variations in those behaviors impact learning.

Chapter 2. Chapter two comprises a study entitled "Spendency: Students' Propensity to Use Game-Based Features." This manuscript is under review at the International Journal of Artificial Intelligence in Education and is authored by Erica L. Snow, Laura K. Varner, G. Tanner Jackson, and Danielle S. McNamara. The abstract from this manuscript is below.

Using students' process data from the game-based Intelligent Tutoring System (ITS) iSTART-ME, the current study examines students' propensity to use system currency to unlock game-based features, (i.e., referred to here as *spendency*). This study examines how spendency impacts students' interaction preferences, in-system performance, and learning outcomes (i.e., self-explanation quality, comprehension). A group of 40 high school students interacted with iSTART-ME as part of an 11-session experiment (pretest, eight training sessions, posttest, and a delayed retention test). Students' spendency was negatively related to the frequency of their use

of personalizable features. In addition, students' spendency was negatively related to their in-system achievements, daily learning outcomes, and performance on a transfer comprehension task, even after factoring out prior ability. The findings from this study indicate that increases in students' spendency are systematically related to their selection choices and may have a negative effect on in-system performance, immediate learning outcomes, and skill transfer outcomes. The results have particular relevance to game-based systems that incorporate currency to unlock features within games as well as to the differential tradeoffs of game features on motivation and learning.

Chapter 3. Chapter three comprises a study entitled "Stealth Profiling of Nuanced and Dynamical Self-Regulative Behaviors." This manuscript is under review at the International Journal of Educational Data and is authored by Erica L. Snow, Aaron D. Likens, Laura K. Varner, and Danielle S. McNamara. The abstract from this manuscript is below.

Self-regulation is a critical skill for students' academic success. However, the predominate use of self-report measures in self-regulation research limits what can be captured about how students regulate their behaviors during learning tasks. The current work utilized a novel approach to capture self-regulative behaviors by employing dynamical analyses to classify students' sequences of choices within an adaptive learning environment. Random walk analyses and Hurst exponents were used to classify students' interaction patterns as random or deterministic. Forty high school students interacted with the game-based system, iSTART-ME, for 11-sessions (pretest, 8 training sessions, posttest, and a delayed retention test). Analyses revealed that students who interacted in a more deterministic pattern also generated higher quality self-explanations during training sessions. The results point toward the potential for

dynamical analyses such as random walk analyses and Hurst exponents to provide stealth assessments of students' self-regulation while learning.

The culmination of these two projects begins to provide evidence regarding the impact that variations in students' interactions with game-based features have on learning outcomes. Although the integration of game-based features within adaptive learning environments has yielded overall positive attitudinal effects (Jackson & McNamara, in press; Rai & Beck; 2012), few studies have investigated how students' regulate their interactions with such features from both a coarse-grained and fine-grained level (a notable exception includes: Sabourin et al., 2012). The combined results presented in this thesis are some of the first to use log data from game-based environments to investigate how users approach various learning tasks.

Chapter 2

“Spendency: Students’ Propensity to Use Game-Based Features”

Submitted to the International Journal of Artificial Intelligence in Education

Game-based features are increasingly popular devices within educational learning environments as a means to increase student interest and promote long-term, persistent interactions within learning systems (Cordova & Lepper, 1996; Jackson, Boonthum, & McNamara, 2009; Rai & Beck, 2012). Games are clearly complex, but some elements include devices such as performance contingent incentives (e.g., currency), personalizable agents (e.g., avatar edits), and navigational choices (e.g., option to turn left or right). The incorporation of such interactive features within learning environments has been shown to increase students’ engagement and motivation in learning tasks (Cordova & Lepper, 1996; Jackson & McNamara, in press; Rai & Beck, 2012).

While educational games and game-based features clearly have some demonstrated benefits, particular in terms of enhancing students’ levels of motivation and engagement, there remains some controversy regarding their benefits to learning. One controversy regards to degree to which game-based features may act as seductive distracters. As such, these features may increase opportunities for students to engage in behaviors that distract from learning goals. For example, adding rewards or personal achievements to educational environments may increase students’ investment with the system (Snow, Jackson, Varner, and McNamara, 2013c), but also afford behaviors that are distracting and tangential to the learning task (e.g., spending time viewing trophies). In general, when students behave in a manner that is divergent and unrelated to the designated task, there can be a negative impact on learning (Carroll, 1963; Rowe, McQuiggan, Robinson, & Lester, 2009; Stallings, 1980). In this study, we further examine these

issues regarding game-based features, focusing on game currency, which is frequently used within games to unlock various features. In particular, we investigate relations among students' propensity to interact with game-based features (spendency), their in-system performance, and learning gains.

Seductive Distracters

Seductive distracters are defined as any irrelevant but engaging stimulus that pulls a student's attention away from the learning goal (Harp & Mayer, 1997). Some examples of seductive distracters include: achievements pop-ups (e.g., level advancement notifications), interface customization (e.g., editing system agents), and irrelevant information embedded within text. Students' interactions with seductive distracters have been shown to have a negative impact on their learning gains (Garner, Alexander, Gillingham, Kulikowich, & Brown, 1991; Godwin & Fisher, 2011; Hidi & Baird, 1986; Harp & Mayer, 1997; Harp & Mayer, 1998; Rowe et al., 2009). For instance, Godwin and Fisher (2011) found that when students spent more time focusing attention on seductive distracters, such as visual representations (i.e., posters, artwork, maps), their overall accuracy on comprehension learning assessments decreased. Similarly, Harp and Mayer (1997) showed that when students were asked to read a passage with an embedded seductive distracter (i.e., irrelevant information), they comprehended the passage at a more shallow level compared to those students who read the passage without the presence of seductive distracters. In sum, seductive distracters potentially disconnect students from the learning goal by decreasing the amount of time and attention a student may spend focusing on a designated task (Harp & Mayer, 1998).

Intelligent Tutoring Systems and Gamification

Intelligent Tutoring Systems (ITSs) are computer-based learning environments that adapt content and feedback to students based on their individual needs (Murray, 1999; VanLehn, 2011). These systems are effective at improving student performance and promoting overall learning gains (Woolf, 2009). When ITSs specialize in complex skill acquisition they may require extensive amounts of training and practice, which in turn, can result in student disengagement and boredom (Bell & McNamara, 2007; Jackson & McNamara, in press; McNamara, Jackson, & Graesser, 2009). This increase in student disengagement over time has led researchers to investigate how they can improve student motivation and engagement within ITSs.

One solution that researchers have found to combat student boredom within educational environments is through gamification. Gamification is the process of adding game-based features or elements (e.g., incentives, editable avatars, and navigational choices) into a non-game based context for the purpose of increasing students' engagement in a task (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). Previous work has revealed that when interactive features are assimilated into learning environments (i.e., gamification), students have shown increased engagement and motivation in tasks (Cordova & Lepper, 1996; Baker, Corbett, Koedinger, & Wagner, 2004; Facer et al., 2004; Jackson & McNamara, 2011; Rowe, Shore, Mott, & Lester, 2010; Sabourin, Rowe, Mott, & Lester, 2011; Snow, Jackson, Varner, & McNamara, 2013a; Rai & Beck, 2012). For instance, work by Rai and Beck found that when game-based features (e. g., personalizable monkey agent) were implemented into a complex learning environment, students reported increased enjoyment. Similarly, Rowe, Shores, Mott and Lester found that as interactive choices within a system increased (i.e., control over where to explore on a map), students' engagement with the system also increased. More recently, Snow and colleagues demonstrated

that students' interactions with personalizable avatar features was negatively related to posttest measures of boredom and positively related to posttest measure of personal control. Thus, this growing body of research has demonstrated that the addition of game-based features within ITSs has been found to have positive effects on students' attitudes.

Although research has demonstrated the benefits of game-based features, these same features may act as seductive distracters that engage students in behaviors that are irrelevant or tangential to the overall learning task, consequently impacting the success of the tutoring system. When researchers have investigated how users' learning outcomes are affected by the presence of seductive game-based features, they have found mixed results (Snow et al., 2013a; Rai & Beck, 2012; Rowe et al., 2009). For instance, Rowe and colleagues examined how students' interactions with seductive game-based features influenced their learning outcomes. They found that embedded navigational choices in an open game-based environment (i.e., Crystal Island), negatively influenced posttest performance. In contrast, Rai and Beck found that game-based features embedded within a math tutor had no impact on students' learned math skills. These contradictory findings indicate that more work needs to be conducted to fully understand how game-based features designed to increase engagement ultimately impact learning outcomes.

One way that researchers may begin to shed light on the influence that these features have on learning is to examine students' process data (Snow, Likens, Jackson, & McNamara, 2013). Indeed, students may choose to interact with game-based features differently, which in turn may influence the impact that these features have on learning. The current study attempts to gain a deeper understanding of this issue by examining variations in students' use of one specific game-based feature, in-system currency. In-game currency has been used in a variety of game-based environments as a way to promote engagement and provide performance incentives

(Gillam, & James, 1999; Jackson & McNamara, 2013; Kim, Park & Baek, 2009; Lee, Goh, Chua, & Ang, 2010; Rymaszewski, 2007). For instance, within the mobile content sharing game Indagator, users earn currency by adding content the digital environment. Users' currency can then be used to customize their experiences and unlock various system features (e.g., new encounters or game-based objects). Within Indagator, currency is also used to indicate users' rank and experience within the game environment (Lee et al., 2010). Although there seems to be a wealth of systems that use some form of currency, relatively few studies have been conducted to examine how this game element may act as a seductive distractor thus, impacting students' performance. The current study is one of the first studies to utilize students' process data to gain a deeper understanding of how varying levels of interactions with in-game currency impact in-system performance and learning outcomes within the context of a game-based system.

iSTART

Interactive Strategy Training for Active Reading and Thinking (iSTART) is an ITS designed to improve students' comprehension of science texts through the instruction of reading strategies (McNamara, Levenstein, & Boonthum, 2004). iSTART provides students with instruction to use reading strategies, including comprehension monitoring, predicting, paraphrasing, elaborating, and bridging, in the context of self-explaining complex texts. Self-explanation, the process of explaining the meaning of text to oneself, has been found to improve performance on a wide range of tasks including problem solving, generating inferences, and deep comprehension of text (Chi, Basskok, Lewis, Reimann, & Glaser, 1989; McNamara, 2004). iSTART combines the process of self-explanation with comprehension strategy instruction so that students engage in deep processing of text and in turn, improve their use of strategies by employing them within self-explanations (McNamara, 2004). Students using iSTART have demonstrated significant

improvements in comprehension, with average Cohen's *d* effect sizes ranging from .68 to 1.12 depending on the learner's prior knowledge (Magliano et al., 2005; McNamara, 2009; McNamara, O'Reilly, Best, & Ozuru, 2006; McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007).

The iSTART system is divided into three training modules: introduction, demonstration, and practice. During the introduction module, students are introduced to three animated agents (two students and one teacher) who discuss the definition of self-explanation and the iSTART reading strategies. The animated agents provide descriptions and examples of each reading strategy, followed by formative assessments intended to measure students' overall understanding of the strategies. After the introduction module is complete, students are transitioned into the demonstration module. In this module, two animated agents (one teacher and one student) apply the various strategies to example texts. Then, students are asked to identify the specific strategies that the agents have employed (see Figure 1). Finally, the practice module of iSTART provides students with an opportunity to apply the strategies they have learned to science texts. In this module, students are given example texts and asked to self-explain various target sentences from the designated text. Then, the teacher agent provides students with formative feedback on their practice self-explanations.



Figure 1. Screen shot of the demonstration module.

In addition to the three training modules, iSTART contains an extended practice module, which operates in the same manner as the practice module. The difference, however, is that extended practice contains a larger number of available texts, which are intended to sustain practice for several months. In addition, teachers can add their own texts that are not currently in the system and assign them to their students.

To provide students with appropriate feedback, an algorithm was developed to assess the quality of students' individual self-explanations. The algorithm utilizes a combination of latent semantic analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007) and word-based measures to assess students' self-explanations, yielding a score from 0 to 3. A score of "0" is provided to self-explanations that are either composed of irrelevant information or are too short, whereas a score of "1" is applied to self-explanations that relate to the sentence itself and do not elaborate upon the provided information. A score of "2" implies that the student's self-explanation incorporates information from the text beyond the target sentence itself. Finally, a score of "3" suggests that a students' self-explanation has incorporated additional information about the text at a global level. This added information may relate to a student's prior knowledge

outside of the target text, or it may focus on the overall theme or purpose of the text. Previous research has shown that the iSTART algorithm provides self-explanation scores that are comparable to human raters (Jackson, Guess, & McNamara, 2010; McNamara, Boonthum, Levinstein, & Millis, 2007).

iSTART-ME

Although the training and extended practice in the iSTART system were shown to improve students' performance across time (Jackson, Boonthum, & McNamara, 2010), the repetitive nature of the extended practice module resulted in some student disengagement and boredom (Bell & McNamara, 2007). To address this problem, the iSTART extended practice module was incorporated into a game-based environment called iSTART-ME (Motivationally-Enhanced; Jackson & McNamara, 2013). Recent work with iSTART-ME has shown that the addition of game-based features has enhanced students' motivation, engagement, and persistence (Jackson & McNamara, 2013; Snow et al., 2013c).



Figure 2. Screen shot of iSTART-ME selection menu

The extended practice interface in iSTART-ME is controlled through a selection menu where students can choose to read and self-explain new texts, personalize characters, play mini-

games, earn points, purchase rewards, and advance levels (see Figure 2). In addition, students can view their current level, the number of points earned, and trophies earned through game play (see Figures 3). Students earn points in the system by interacting with texts, either within the context of generative games or a non-game practice module (see Jackson & McNamara, 2013 for more detail). Students earn trophies in the system based on their performance in the various practice games. These trophies are linked to the number of points that students earn in each game (see Jackson & McNamara, 2013, for more detail). As students collect more points in the system, they subsequently progress through a series of levels ranging from 0 to 25, where new game-based features are unlocked at the different levels. Each level requires more points to proceed than the previous level. Thus, students must exert more effort as they progress through the higher levels in the system.

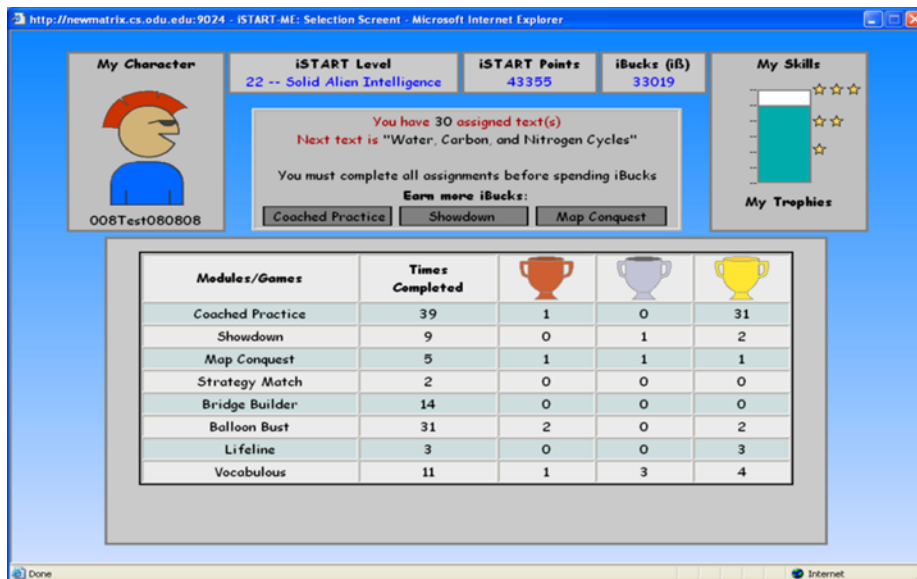


Figure 3. Screen shot of trophy and level view on iSTART-ME selection menu

The points that students earn throughout their interactions with iSTART serve as a form of currency (iBucks) and can be used to unlock game-based features within the system. Students' earned iBucks map on to the number of points that they have earned within the system. For

example, if Jane earned 500 points while playing a practice game, she also has earned 500 iBucks that will automatically appear on the game-based menu. There are two primary uses for iBucks: personalization of system features and playing mini-games. As students choose to interact with these features, the appropriate number of iBucks will be deducted from their total; however, no points are deducted. Each of these system features requires either 300 or 450 iBucks (see Table 1 for a list of the potential uses for the iBucks). The game-based mechanic of currency was embedded within the system as a means to engage users' interests and provide them with a sense of agency over their learning paths (Jackson & McNamara, 2013).

Table 1.

Cost of interactions for every game-based feature

Personalizable Interactions	Cost in iBucks
Avatar Edits	300
Background Theme Edits	300
Pedagogical Agent Change	300
Practice Interactions	Cost in iBucks
Strategy Match	300
Bridge Builder	300
Vocabulous	300
SE Lifeline	300
Balloon Bust	300
Dungeon Escape	450

One potential use of the iBucks is for the personalization of the system environment. These options were implemented into the system as a means of increasing students' investment in the system, as well as provide them with a sense of control over their learning environment. Within the extended practice interface, students have the option of changing numerous features, such as background colors or teacher agents. For instance, students can use their iBucks (iSTART points) to purchase new hairstyles or headgear for their avatars (see Figure 4 for examples).

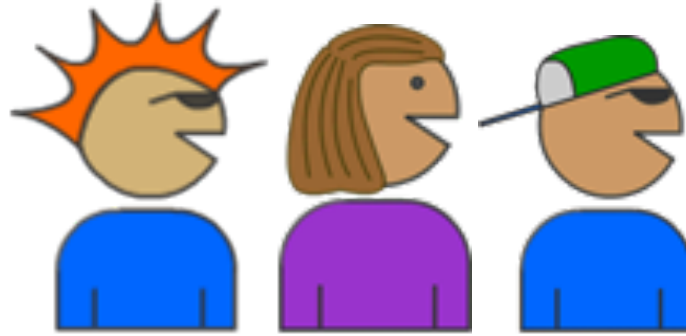


Figure 4. Example avatar configurations

In addition to personalizable interactions, students can use their iBucks to interact with practice features (mini-games). These practice features were added to iSTART-ME to provide students with opportunities to practice identifying the various self-explanation strategies. For instance, in the mini-game, Balloon Bust, students are presented with a target sentence and an example of a self-explanation. Students must then decide which strategy was used to generate the self-explanation and click the corresponding balloons on the screen (see Figure 5).

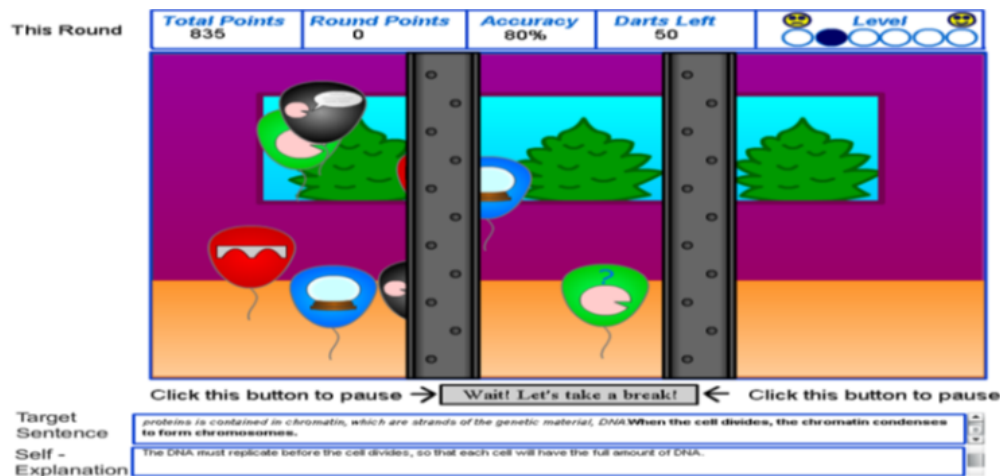


Figure 5. Screen shot of Balloon Bust

Current Study

Previous work has provided some insight into the impact that game-based features have on student learning. However, there remain questions that need to be addressed to understand the

influence that students' interactions with these features may have on in-system performance and learning outcomes. The current study attempts to address this issue by examining students' process data to examine how they interact with in-game currency (i.e., spendency) and investigate four primary questions.

- 1) How does students' spendency vary as a function of individual differences in prior skill level and self-reported attitudes?
- 2) How do students' interaction patterns with various game-based features vary as a function of their spendency?
- 3) How do differences in students' spendency impact their in-system performance?
- 4) How do differences in students' spendency impact learning outcomes?

One contribution of this study stems primarily from the analysis of process data to investigate users' propensity to spend in-game currency as a way to interact with game-based features (spendency), such as practice and personalizable features. Examining the impact of variations in students' interactions further explores the impact that game-based features have on performance and learning gains. Results from this study may begin to provide researchers with deeper understanding of how gamification can influence learning goals within game-based systems.

METHOD

Subjects

Participants in this study included 40 high-school students from a mid-south urban environment (50% male; 73% African American, 17% Caucasian, 10% other nationalities; average grade level of 10.4; average age = 15.5 years). The sample included in the current work is a subset of 124 students who participated in a larger 11-session study that compared three conditions: iSTART-

ME, iSTART-Regular, and no-tutoring control. In that study, both iSTART and iSTART-ME were shown to improve students' strategy performance from pretest to posttest (Jackson & McNamara, 2013). The results presented in the current study solely focus on the students who were assigned to the iSTART-ME condition, as they were the only students who had access to the full game-based selection menu.

Procedure

All students completed the full 11-session experiment that consisted of a pretest, eight training sessions, a posttest, and a delayed retention test. During the first session, participants completed a pretest survey that included measures of motivation, attitudes toward technology, prior self-explanation (SE) ability, and prior reading comprehension ability. During the following eight sessions, participants took part in the training portion of the experiment, each lasting an hour. In these sessions, students interacted with the full game-based menu, where they had access to mini-games, texts, and interactive features. All students in the current study interacted with the iSTART-ME system for all eight training sessions. After the eight training sessions, students completed a posttest that included similar measures to the pretest. One week after the posttest, students completed a retention test that contained measures similar to the pretest and posttest (i.e., self-explanation ability).

Measures

Strategy performance. Self-explanation (SE) quality was measured during pretest, posttest, and at retention by asking students to read through a text one sentence at a time and generate a self-explanation when prompted. Training strategy performance was assessed during sessions two through eight by measuring the self-explanations students produced while engaging in the generative practice games. All generated self-explanations were scored using the

automated iSTART assessment algorithm (McNamara et al., 2007; Jackson, Guess, & McNamara, 2010).

Reading comprehension ability. Students’ reading comprehension ability was assessed at pretest using the Gates-MacGinitie Reading Comprehension Test (MacGinitie & MacGinitie, 1989). This test includes 48 questions designed to assess the general reading ability of each student by asking them to read a passage and then answer comprehension questions about the passage. This test is a well-established measure of student reading comprehension, which provides researchers with in-depth information about students’ starting literacy abilities ($\alpha=.85-.92$, Phillips, Norris, Osmond, & Maynard, 2002).

Attitudes and motivation. During each training session, students were asked to answer a battery of questions that assessed their daily enjoyment and motivation during their time interacting with iSTART-ME (see Table 2 for daily questions). These questions have been used in previous studies to assess individuals’ motivation and enjoyment within adaptive systems (Jackson, Graesser, & McNamara, 2009; Jackson & McNamara, in press; Snow, Jackson, Varner, & McNamara, 2013b). Students’ were asked to rate their experience within the system eight times (once per session). These scores were combined and averaged to create a mean enjoyment and motivation score.

Table 2.

Daily enjoyment and motivation measures

Dependent Measure	Response Statement	Response Scale
Enjoyment	“I enjoyed my most recent session”	1 - 6
Motivation	“I was motivated to participate in my most recent session”	1 - 6

1 (Strongly Disagree) to 6 (Strongly Agree)

In-system performance. Students' in-system performance was measured through the highest level that the student achieved throughout iSTART-ME training. As described earlier, students advance to higher levels within the system by earning performance points within practice games and mini-games. There are 25 levels and each level requires more points to proceed than the previous level. This measure is reflective of students' daily strategy performance within the context of the iSTART-ME system.

Learning transfer. iSTART-ME was designed to teach students strategies to improve their reading comprehension. Learning outcomes were measured through an open-ended comprehension measure. In the current study, the transfer of self-explanation training to reading comprehension was assessed at the retention test using a science passage-specific comprehension measure. In this task, each student was asked to read an assigned science text. After they finished reading, students were then presented with a series of open-ended questions that they were asked to answer based on their recollection of the text. These questions were designed to assess low-level and deep-level text comprehension. Two raters independently scored these open-ended questions. Initial inter-rater agreement was high, with an overall kappa of .951.

Spendency. Students' propensity to use system currency as a way to engage with the game-based features was defined as their spendency. This measure was calculated by dividing the total number of iBucks students spent on mini-games and personalizable features by the total iBucks each student earned within the system (total points spent / total points earned). This proportion yielded each student's spendency (tendency to spend) within iSTART-ME. This measure allows us to examine the degree to which students spent their in-system currency on various game-based features (with respect to the number of iBucks earned). This measure allows

us to investigate how much of their available resources (i.e., iBucks) a student used while in iSTART-ME.

Interaction Patterns. All students' interactions in iSTART-ME were logged and recorded within the iSTART-ME database. This process data was then organized according to the function afforded by each feature: generative practice, identification mini-games, personalization of interface features, and screens to monitor system achievements and progress (see Jackson & McNamara, 2013, for detailed descriptions). Through the use of a statistical sequencing procedure similar to that used in D'Mello, Taylor, and Graesser (2007) and this time stamped data set, we calculated the probability of each student's set of interactions within the system. This calculation can be described as $L[It \rightarrow Xt+1]$, where, put simply, we calculated the probability of a student's next interaction (X) with an interface feature given their previous interaction (I). This sequencing procedure allows us to trace students' interaction trajectories across time.

RESULTS

Spendency

To characterize how students spent their earned iBucks, spendency was calculated using a proportional formula (total points spent / total points earned). Within the current study, students' spendency varied considerably (range =0.0 to .99, $M=0.49$, $SD=0.26$).

Individual Differences and Spendency

To gain a deeper understanding of how individual differences related to students' tendency to spend iBucks, Pearson correlation analyses were conducted on pretest reading and self-explanation scores. Results from these analyses revealed no significant relation between students' spendency and their prior reading ability ($r= -.21$, $p=.19$) or self-explanation

performance ($r = -.10$, $p = .55$). Thus, these results indicate that individuals' reading ability and strategy performance did not influence their propensity to spend iBucks on interactions with game-based features.

Motivation, Enjoyment and Spendency

To assess how spendency related to students' self-reported enjoyment and motivation, Pearson correlation analyses were conducted using students' mean enjoyment and motivation scores. Results from these analyses revealed no significant relation between students' spendency and their self-reported enjoyment ($r = -.18$, $p = .26$) or motivation ($r = -.21$, $p = .19$). Thus, students' propensity to spend their earned iBucks does not seem to be the result of boredom or disinterest. Overall, students tended to report high ratings of enjoyment ($M = 5.0$, $SD = .75$) and motivation ($M = 5.37$, $SD = .80$) within the system. These results are similar to recent work showing that students generally rate their interactions with educational games as positive experiences (Jackson & McNamara, 2013; Rodrigo & Baker, 2011).

Interaction Patterns

The current paper examines the extent to which students' tendency to spend in-game currency relates to their interactions patterns within the iSTART-ME system. Using the aforementioned probability analysis, we can examine students' transitions between and within the four categories of game-based features. To investigate how students' spendency related to their custom system trajectories, Pearson correlations were calculated using students' spendency levels and their transitional probability patterns (see Figure 6 for interaction patterns). Figure 6 provides a visual display of the relation between spendency and interactions patterns, with numbers inside a box representing the correlation between spendency and students' selecting the same feature again. For instance, in the identification mini-game box, we see the value $r = .05$. This value indicates

that spendency was not significantly related to students' choice in selecting an identification mini-game after they had just completed an identification mini-game. The numbers near a transition line indicate the relation between spendency and students' choice to transition from one feature to another.

Overall, results from this analysis demonstrated a significant positive relation between spendency and two probability interactions. First, students who had a higher spendency were more likely to interact with a personalizable feature after playing a generative practice game ($r=.34, p<.05$). Similarly, students with a high spendency were also more likely to return to a generative practice game after interacting with a personalizable feature ($r=.41, p<.01$).

Interestingly, these results reveal that students who spent a higher proportion of their iBucks were not just jumping between features that did not reinforce learning strategies. Indeed, these students seemed to be interacting in a pattern where they practiced generating self-explanations and then moved to personalizing various game-based features.

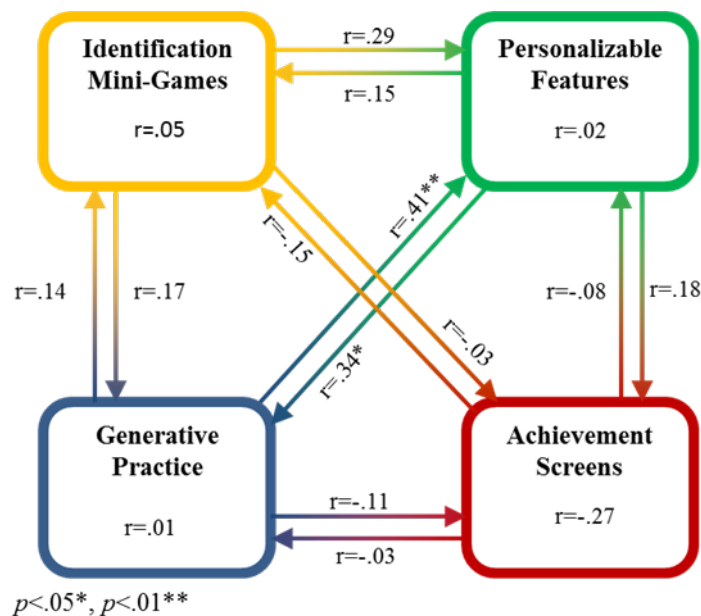


Figure 6. Relation between interaction patterns and spendency

In-System Performance

Within iSTART, students advance to higher achievement levels based on their performance in the system. To investigate how spendency impacted achievement levels, we used a hierarchical linear regression model to factor out students' pretest strategy knowledge. In Model 1 of this analysis, we included pretest strategy knowledge scores to predict achievement levels. Results from this analysis revealed that pretest strategy knowledge was a significant predictor of students' achievement levels in the system ($R^2=.36$, $F(1,38)=21.04$, $p<.01$; see Table 3). In Model 2, we examined how spendency predicted achievement levels over and above pretest strategy knowledge. Results from this analysis indicated that spendency was a significant (negative) predictor of student achievement levels over and above prior strategy knowledge (R^2 change=.17, $F(1,37)=20.77$, $p<.01$; see Table 3). This analysis revealed that students' spendency accounted for 17% of the variance in students' in-system achievement levels over and above their prior strategy knowledge, in particular indicating lower performance as a function of greater spendency.

Table 3.

Hierarchical linear regression analyses predicting achievement level

Variable	B	SE	β	ΔR^2
Model 1				.36**
Pretest SE	5.75	1.25	.60**	
Model 2				.17 **
Pretest SE	5.36	1.10	.56**	
Spendency	-9.41	2.56	- .42**	

$p < .01^{**}$

Strategy Performance

We further examined how variations in students' spendency related to their daily strategy performance. Pearson correlations were conducted (see Table 4) to investigate relations between students' spendency and their daily self-explanation scores. Results from this analysis indicated

that there was a negative correlation between students' spendency and self-explanation quality on days 3, 4, 5, 6, 7, and 8. However, there was no significant relation between spendency and self-explanation quality on days 1 and 2 of training. This non-significant relation is not unsurprising because most students are just beginning to accumulate iBucks on days 1 and 2.

Table 4.

Spendency and Daily Strategy Performance

Daily Strategy Performance	Spendency
Session 1	-.147
Session 2	-.238
Session 3	-.392*
Session 4	-.448**
Session 5	-.479**
Session 6	-.354**
Session 7	-.416**
Session 8	-.391**

p<.05*, p<.01**

To further examine these relations, we conducted separate hierarchal regression analyses on students' self-explanation quality scores for each of the six significant training days in Table 4. These analyses investigated how students' spendency predicted self-explanation scores over and above prior self-explanation ability (i.e., self-explanation scores at pretest). This is reflected by the R² change attributable to the variance accounted for by spendency after entering prior strategy ability (i.e., pretest self-explanation scores) in the regression model (see Table 5). These analyses revealed significant models and R² change values for session 3, $F(1,37)=6.79, p<.05, R^2=.31, R^2\text{Change}=.13$ (i.e., see session 3 in Table 5), session 4, $F(1,37)=9.13, p<.01, R^2=.29, R^2\text{change}=.18$, session 5, $F(1,37)=11.52, p<.01, R^2=.38, R^2\text{change}=.20$, session 6, $F(1,37)=4.97, p<.05, R^2=.17, R^2\text{change}=.11$, session 7, $F(1,37)=7.31, p<.01, R^2=.25, R^2\text{change}=.15$ and session 8, $F(1,37)=6.18, p<.05, R^2=.30, R^2\text{ change}=.15$.

Finally, the current study investigated how students' interactions with game-based features impacted their long-term learning outcomes (i.e., self-explanation quality). Pearson correlations were calculated to examine how spendency related to posttest, and retention self-explanation scores. This analysis revealed that spendency was not related to self-explanation quality at posttest ($r = -.195, p = .227$) or retention ($r = -.231, p = .187$). Taken together, these results indicate that spendency had an immediate negative effect on strategy performance, but it did not appear to be a detriment over time.

Table 5.

Hierarchical Linear Regressions Predicting Self-explanation Quality Spendency and Prior Strategy Ability

Self-Explanation Quality	β	ΔR^2	R^2
Session 3			.31*
Prior Strategy Ability	.40	.19**	
Spendency	-.36	.13*	
Session 4			.29**
Prior Strategy Ability	.29	.11*	
Spendency	-.42	.18*	
Session 5			.37**
Prior Strategy Ability	.37	.17**	
Spendency	-.45	.20**	
Session 6			.17*
Prior Strategy Ability	.20	.05	
Spendency	-.34	.12*	
Session 7			.25*
Prior Strategy Ability	.27	.09	
Spendency	-.39	.16**	
Session 8			.30*
Prior Strategy Ability	.38	.14*	
Spendency	-.39	.16*	

p<.05 *, p<.01 **

Learning Transfer

To investigate how students' spendency related to transfer learning over and above prior ability, we conducted a hierarchical linear regression. In Model 1 of this analysis, we included pretest self-explanation scores to predict reading comprehension ability. Results from this analysis revealed that pretest strategy performance was a significant predictor of students' scores on the reading

comprehension transfer task ($R^2=.10$, $F(1,38)=4.31$, $p<.05$; see Table 6). In Model 2, we examined how spendency predicted scores on the comprehension transfer task over and above pretest strategy knowledge. Results from this analysis indicated that spendency was a significant predictor of students' comprehension scores over and above prior strategy knowledge (R^2 change=.13, $F(1,37)=6.16$, $p<.05$; see Table 6).

Table 6.

Hierarchical linear regression analyses predicting reading comprehension scores on the transfer task

Variable	B	SE	β	ΔR^2
Model 1				.10*
Pretest SE	1.27	.61	.32*	
Model 2				.13*
Pretest SE	1.14	.58	.29	
Spendency	-3.40	1.37	-.36*	

*p < .05**

DISCUSSION

The current study adds to the literature by using process data to investigate how students' propensity to use in-system resources (i.e., iBucks) to engage with game-based features (spendency) impacted their interaction trajectories, in-system performance, attitudes, and learning gains. We introduce the term spendency as a term to describe students' use of in-game currency. The importance of this game feature is not isolated solely to the iSTART-ME system. Of course the term spendency can be applied to any system that utilizes a gamification technique where in-game currency is applicable. As the gamification of ITSS becomes more prevalent, it is important to understand how users choose to engage with various types of game-based features and the impact of those interactions.

Initial results from this study revealed that students' spendency was related to a specific interaction pattern within the iSTART-ME system. Specifically, when students' had a high

spendency they were more likely to engage in an interaction loop between generative practice games and personalizable features. Thus, these students played a generative practice game and then chose to customize the system in some way. Once they had customized the system, these students tended to revert back to playing a generative practice game. These trajectories could be attributable to a couple of factors. The first is that the personalizable features may be the most visually salient features on the interface (see Figure 2), thus, directing students' attention and resources towards them more often. The second is that these features are the only elements within the system that afford detachment from educational content. High spendency students may have interacted with these features as a way to get a mental break from the strategy instruction embedded within the system. Interestingly, attitudinal results from the current study indicate that spendency was not related to motivation or engagement. Thus, these students did not appear to simply be interacting with the personalizable features out of boredom.

Current literature indicates that interactions with game-based features can have a negative impact on students' learning outcomes (Rowe et al., 2009). Results from the current study add a deeper understanding of the impact of game-based features by revealing that students' spendency had an immediate negative impact on in-system performance, daily strategy performance, and learning transfer. Thus, students who were more interested in spending their earned currency did not perform well during training and also had lower scores on the learned skill transfer tasks. This finding could be due to students placing a higher importance on spending their earned resources to interact with features and, therefore, spending less time engaged in the learning tasks. Overall, these results support the hypothesis that overexposure to game-based features may act as seductive distracters that pull students' attention away from the designated task thus, negatively influencing their ability to transfer learned skills to new tasks.

ITS researchers have used numerous types of game-based features to leverage students' enjoyment and (indirectly) impact learning. However, the influence of these features is not completely understood. The results from the current study suggest that interactions with game-based features can vary; however, when students put a high amount of importance on these features there may be immediate and long-term negative consequences on task performance. This finding is important because it shows that the consequences of embedded game-based features may primarily impact immediate learning outcomes and students' ability to transfer their learned skills to new tasks. These results underscore the importance of understanding of both the immediate and long-term effects of game-based features that are integrated into learning environments. Although previous work has shown positive attitudinal effects from students' exposure to game-based features (Jackson & McNamara, in press; Snow et al., 2013a; Rai & Beck, 2012), the current results suggest that there are some potential consequences, at least for immediate and transfer performance. As learning environments are developed, the addition of game-based features may or may not be appropriate depending on the learning task. For instance, if the learning goal of a system is to show immediate gains in domain knowledge, the inclusion of game-based features may be detrimental to that goal. However, in a system such as iSTART-ME, which was designed to engage students in practice over long periods of time, the immediate effects of game-based features may not be as important in the long-term. These initial results should warn developers to ensure that the inclusion of features does not interfere with the learning goal relative to the timeframe of the system.

Although it is important to understand the impact of game-based features on system and task performance, it is also important to identify students who are more inclined to engage with these features. Understanding the individual differences that drive students' interactions within

systems may help researchers develop environments wherein traits of the interface adapt depending on the users' affect. The current study took steps to identify individual differences that may influence students' propensity to interact with game-based features. However, results indicate that spendency was not related to students' scores on the Gates-MacGinitie Reading Comprehension Test or their prior strategy performance. These results suggest that both high and low ability students use their in-game currency to interact with game-based features at an equivalent rate. These preliminary results begin to show that both high and low ability students engage with seductive distracters, though work remains to further elucidate how attitudinal measure influence whether or not students tend to attend to seductive distracters or ignore them.

The analyses presented in the current work are intended as a seed for future studies by providing evidence that system log-data can provide valuable metrics for the way in which users behave with game-based environments. Going forward, we plan to include these initial behavioral measures in a student model within the iSTART-ME system. Such analyses will be especially valuable if the iSTART-ME system is able to recognize non-optimal learning behaviors and steer students toward more effective behaviors. For instance, if a student is engaging in a high spendency behaviors, it may be beneficial for iSTART-ME to have the capability to recognize this trend and prompt the student toward less spending and more generative practice.

Another future direction of study regards finding a "sweet spot" or balance between game-based and pedagogical features, successfully promoting both engagement and learning. Results from this study and others (Jackson & McNamara, 2013) suggest that exposure to game-based features may negatively influence immediate performance and learning outcomes. However, other research has shown that game-based features have a positive effect on students'

attitudes (Rai & Beck, 2012; Snow et al., 2013c). Combined, these findings suggest that future work should further focus on the complex interplay between disengagement and learning within game-based environments.

CONCLUSION

Game-based features have been incorporated within a number of adaptive environments, principally in an attempt to enhance students' engagement and interest in particular learning tasks. While this movement toward more engaging and creative learning environments is compelling, it is important for researchers and educators to more fully understand the impact that these features may or may not have on students. Our study adopts a novel approach to achieving that objective by using process data to investigate how students' spendency is related to training performance, immediate learning outcomes, and learning transfer. Previous work has found that game-based features may act as seductive distracters, negatively impacting learning outcomes (e.g., Harp & Mayer, 1997). Such work may lead to the conclusion that game-based features and games should not be incorporated within learning environments. Our overarching assumption is that the impact of game-based features can have a negative impact on learning, thus, they should be incorporated into systems with caution. The results of this study further support that assumption. Specifically, outcomes from current work indicate that students' propensity to spend earned iBucks to interact with embedded game-based features led to immediate negative consequences during training performance and at transfer. Thus, initial results reveal that this specific gamification technique (i.e., currency) has the potential divert students' attention and negatively impact their target skill acquisition within a game-based environment.

Chapter 3

“Stealth Profiling of Nuanced and Dynamical Self-Regulative Behaviors”

Submitted to the Journal of Educational Data Mining

Students’ ability to regulate, control, and monitor their behaviors while learning is considered crucial to academic motivation and success (Schunk & Zimmerman, 2003). Indeed, a large body of research has focused on the behaviors, cognitive states, and affect associated with self-regulated learning (SRL; Butler & Winne, 1995; Harris, Friedlander, Sadler, Frizelle, & Graham, 2005; Jarvela & Jarvenoja, 2011; McCombs, 1989; Pintrich & De Groot, 1990; Schunk, 2008; Winters, Greene, & Costich, 2008; Winne, 2013; Winne & Hadwin, 2013; Wolters, 2011; Zimmerman, 1990, 2008; Zimmerman & Schunk, 1989, 2001, 2013). Collectively, this research has established a relationship between various assessments of self-regulated learning (SRL) and a host of positive learning strategies and outcomes, including task performance in laboratory studies as well as academic performance. An overarching assumption across this research is that learning outcomes are enhanced when students more accurately monitor their own progress and performance (e.g., Harris et al., 2005; McCombs, 1989) and periodically self-assess their learning development and outcomes (e.g., Ariel, Dunlosky, & Bailey, 2009; De Bruin, Thiede, & Camp, 2001).

While informative, self-regulation research has been somewhat hamstrung by the traditional use of self-report measures to assess students’ behaviors and intentions during learning tasks. SRL is generally assessed using self-report questions related to learning strategies, affect toward learning, attitudes in the moment, and so on. A concern arises from the SRL literature that these self-report measures may not fully or adequately capture their target construct (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007; McNamara, 2011; McNamara & Magliano, 2009; Zhou, 2013). An overarching concern regards the frequent mismatch between

students' reports of what they do and observations of their actual performance (McNamara, 2011).

The mismatch between self-reports and behavior may arise from a number of factors. First, self-report relies on the student's memory for past events and behaviors, and these memories can be inconsistent and unreliable. Second, the student may lack a clear understanding what comprises good and poor performance, leading to over or under estimations of SRL. Third, the behaviors, cognitive states, and affect associated with SRL can be difficult to observe because they are often not verbal in nature, and thus, the student may not be conscious of these behaviors, and those behaviors also may not be evident to an observer. Finally, and perhaps foremost, self-regulative and metacognitive strategies are dynamic (Hadwin et al., 2007; Lord, Diefendorff, Schmidt, & Hall, 2010; McNamara & Magliano, 2009). Students often behave and learn differently depending on the domain, context, and task. Learning behaviors dynamically fluctuate between contexts and tasks; and they also fluctuate within tasks as comprehension and learning develops and changes over time. Hence, relatively static measures of SRL may not adequately capture nuanced changes in how learners modulate their behaviors across varying goals and task demands.

Online Measures

Online measures of performance offer an alternative means of capturing the dynamic nature of metacognition and SRL (Hadwin et al., 2007; McNamara & Magliano, 2009; Ventura & Shute, 2013; Winne & Hadwin, 2013; Witherspoon, Azevedo, & D'Mello, 2008). In contrast to *offline* self-report and post assessments, online measures elicit behaviors from the learner, and performance on those tasks is used to inform estimates of SRL. Online measures may be particularly informative of SRL because they can focus on nuanced patterns in students'

behaviors and choice pattern during learning and thus, may be more likely to capture how students exert agency while engaging in learning tasks.

Automated learning environments potentially offer a novel venue for the use of online SRL measures. Such environments open up opportunities to incorporate *stealth* assessments that are embedded within tasks and essentially “invisible” to students (Shute, Ventura, Bauer, & Zapata-Rivera, 2009; Shute, 2011). One approach to stealth assessment is to embed assessments within the learning environment, and collect various performance measures while students interact with the system. Another, and the one we evaluate in this study, is the use of log-data, also referred to as keystroke, mouse click, click stream, or telemetry data (depending on the context). Log-data is essentially the recording of all of a user’s interactions or keystrokes while interacting with an automated system. Notably, the collection of log-data is not *built into* all computerized systems, but rather must be intentionally programmed. When it is collected, log-data can provide a wealth of information, particularly concerning students’ choices and *agency* while engaged with a system.

For instance, Hadwin and colleagues (2007) utilized users’ log-data from the gStudy system to create profiles of students’ self-regulatory behaviors. The gStudy system is a web-based platform designed to investigate students’ annotation (e.g., highlight, label, or classify) of educational content. Hadwin et al. examined how students’ patterns of annotation and study habits helped inform profiles of SRL. They demonstrated that log-data informed profiles of self-regulated behaviors by revealing fine-grained behavioral patterns that students exhibited while studying. Hadwin and colleagues argue that these nuanced patterns would have been missed by self-report measures alone.

Similarly, Sabourin and colleagues (2012) examined how log-data from the narrative-centered environment, Crystal Island, was indicative of students' SRL strategy use (e.g., self-monitoring and goal setting). Sabourin et al. investigated how students' behaviors during game-play (e.g., use of notes, books, or in-game tests) and pretest self-report measures of affect and prior knowledge combined to predict students' level (i.e., low, medium, or high) of SRL strategy use. They found that the inclusion of system log-data significantly contributed to the accurate classification of students' use of SRL strategies. Such research demonstrates that log-data extracted from adaptive environments yield an unobtrusive opportunity to examine the nuanced ways in which students regulate their behaviors during learning tasks.

Dynamical Analysis

In conjunction with online data, dynamical systems theory and its associated analysis techniques offer researchers a means of characterizing patterns that emerge from students' behaviors within an adaptive system. Such an approach treats time as a critical variable in addressing patterns of stability and change. Dynamical analyses focus on the complex and sometimes fluid interactions that occur within a given environment rather than treating behavior as static (i.e., unchanging), as is customary in many statistical approaches.

Dynamical methodologies have been utilized in adaptive systems to investigate the complex patterns that emerge in students' behaviors (Hadwin et al., 2007; Snow, Likens, Jackson, & McNamara, 2013; Soller, & Lesgold, 2003; Zhou, 2013). For example, Snow et al. (2013) used random walk algorithms to visualize how individual differences influenced students' trajectories within a game-based environment. Results from that study revealed that students' trajectories within a game-based environment varied as a function of individual differences in students' reading comprehension ability. Snow and colleagues argue that these nuanced choice

patterns that manifested within students' log-data would have been overlooked using more traditional (e.g., static) statistical analyses.

Such research affords scientists a dynamical perspective of students' behaviors within adaptive environments; however, it reveals little information about how students regulate or control their choices. The current work utilizes two dynamical methodologies, random walks and Hurst exponents, to visualize and classify how patterns in students' behaviors manifest over time and relate to learning gains. Random walks are mathematical tools that provide a graphical representation of a path or trajectory (Benhamou & Bovet, 1989). Thus, random walks afford researchers the opportunity to visualize fine-grained patterns that form in categorical data across time. This technique has been used in a variety of domains, such as economics (Nelson & Plosser, 1982), ecology (Benhamou & Bovet, 1989), psychology (Collins & De Luca, 1993), and genetics (Lobry, 1996). For instance, geneticists have utilized these visualization tools to examine distinct patterns of disease and coupling in gene sequences (Arneodo et al., 1995; Lobry, 1996). More recently, learning scientists have utilized this technique to visualize how interaction trajectories within an adaptive system vary as a function of individual differences (Snow et al., 2013).

While random walk analyses generate visualizations of patterns across time, Hurst exponents (Hurst, 1951) classify the tendency of those patterns. Hurst exponents characterize statistical changes in time series by revealing persistent, random, and antipersistent behavioral trends (Mandelbrot, 1982). When fluctuations in patterns are positively correlated from one moment to the next, they are exhibiting a *persistent* quality. Time series fluctuations exhibiting persistence are assumed to reflect self-organized and controlled processes (Van Orden, Holden, & Turvey, 2003). By contrast, when each moment in a time series is independent of every other

moment, the fluctuations in the times series are exhibiting *random* characteristics. Time series that exhibit random processes reflect a breakdown in system functioning and control (e.g., Peng et al., 1995). Finally, when time series fluctuations are negatively correlated from one moment to the next, they are exhibiting *antipersistent* behavior (Collins & De Luca, 1993). Time series fluctuations exhibiting antipersistent behaviors are assumed to be demonstrating corrective processes (Collins & De Luca, 1993).

The goal of the current study is to investigate self-regulation by combining random walk and Hurst analyses to characterize nuanced behavior patterns that emerge within an adaptive system. Notably, this study does not include self-report surveys found in traditional analysis of students' propensity to self-regulate. Consequently, the objective of this study is not to validate the use of dynamical analyses and log-data using concurrent validity measures. Instead, our objective is to assess the potential of these analyses through convergent validity assessed by examining the link between students' controlled patterns of behaviors and their daily learning outcomes. Organized and controlled behaviors are often associated with students' ability to self-regulate (Butler & Winne, 1995; Ellis & Zimmerman, 2001; Zimmerman, 1990). The current study uses a novel application of random walks and Hurst exponents to capture the nuanced and controlled behavior patterns that manifest within students' log-data collected across multiple sessions within a complex learning environment. Ultimately, dynamical techniques may serve as novel forms of stealth assessment, examining students' propensity to control and regulate their behaviors across time, but without relying on obtrusive survey methodologies.

iSTART-ME

The context of the current study is the game-based learning environment, iSTART-ME (Interactive Strategy Training for Active Reading and Thinking-Motivationally-Enhanced;

Jackson & McNamara, 2013). This system is effective at providing students with instruction on the use of self-explanation and comprehension strategies (Jackson, Boonthum, & McNamara, 2012; Jackson & McNamara, 2013; McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007). iSTART-ME is ideal for the current study because it requires multiple sessions to complete; it includes multiple modules; and students choose their individual paths within the environment. Hence, it affords an environment in which students have agency over their learning paths and objectives.

iSTART-ME is based on a traditional intelligent tutoring system, iSTART (McNamara, Levinstein, & Boonthum, 2004; McNamara, et al., 2007) but integrates games and game-based features to enhance students' motivation, engagement, and persistence over time (Jackson, Boonthum, & McNamara, 2009; Jackson, Dempsey, & McNamara, 2010; Jackson & McNamara, 2013). The game-based features in iSTART-ME were incorporated within iSTART following research emphasizing the importance of factors related to motivation such as students' self-efficacy, engagement, self-regulation, and interest (Alexander, Murphy, Woods, Duhon, & Parker, 1997; Bandura, 2000; McNamara, Jackson, & Graesser, 2009; Pajares, 1996; Pintrich, 2000; Zimmerman & Schunk, 2001). The inclusion of these game-based features has been shown to enhance students' engagement and motivation across multiple training sessions (Jackson & McNamara, 2013).

Both iSTART and iSTART-ME introduce, demonstrate, and provide students with practice using self-explanation reading strategies for complex science texts. This is accomplished in three separate modules referred to as introduction, demonstration and practice (see Jackson, Boonthum, & McNamara, 2009; McNamara et al., 2007). The game-based practice within iSTART-ME is referred to as extended practice. In this interface, students can choose to read and

self-explain new texts, personalize characters, play mini-games, earn points, purchase rewards, and advance levels through the use of an embedded selection menu (see Figure 1). Additionally, within this selection menu, students can view their current level and the number of skill points and trophies earned.



Figure 1. Screenshot of iSTART-ME selection menu

In the extended practice interface, students can choose to generate their own self-explanations within three different practice environments: Coached Practice, Map Conquest, and Showdown. These environments afford students the opportunity to engage in strategy practice and receive feedback on the quality of their self-explanations. Coached Practice is a non-game based method of practice adapted from the original iSTART system. In this environment, a pedagogical agent guides practice and provides students with formative feedback on their generated self-explanations. In contrast, Showdown and Map Conquest are both game-based practice environments. In Showdown, students compete against a computer player by generating self-explanations in an attempt to win points. In Map Conquest, students generate self-explanations to earn dice, which are used to conquer squares on a map (see Figure 2). As students engage with texts in these practice environments, they can earn points that allow them to progress through a series of levels ranging from 0 to 25. Each level requires more points to

proceed than the previous level; thus, students must exert more effort as they advance to higher levels in the system.

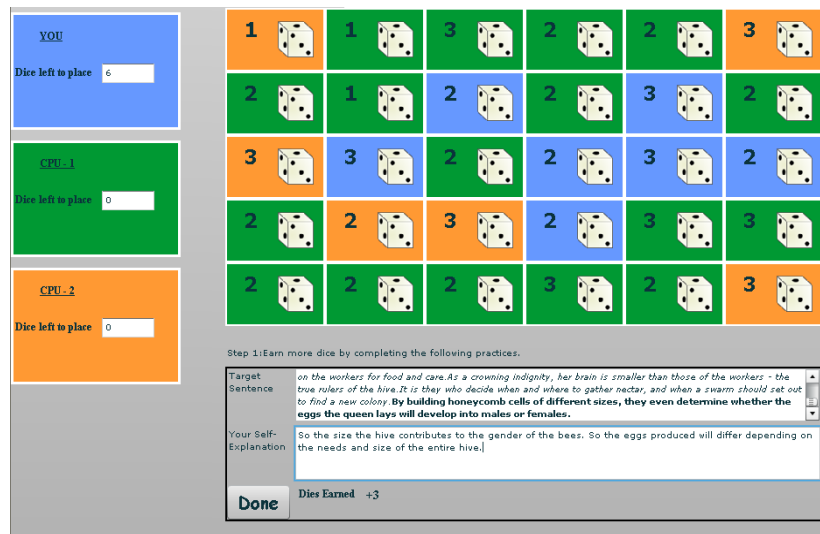


Figure 2. Screenshot of Map Conquest

Students' points also serve as a form of currency (iBucks) that can be used to unlock game-based features within the system. There are two primary uses for iBucks: interacting with personalizable features and playing identification mini-games. Personalizable features were implemented into the system as a means to enhance students' personal investment and sense of control over their learning environment. Within iSTART-ME, students have three personalizable feature options: changing the background theme, customizing an avatar, and editing a pedagogical buddy. Students can also use their iBucks to interact with identification mini-games. These mini-games were added to iSTART-ME to provide students with opportunities to practice identifying the various self-explanation strategies. For instance, in the mini-game, Balloon Bust, students are shown a target sentence and an example of a self-explanation. They must then decide which previously learned strategy was used to generate the example self-explanation and pop (by clicking with the computer mouse) the corresponding balloons on the screen (see Figure 3).

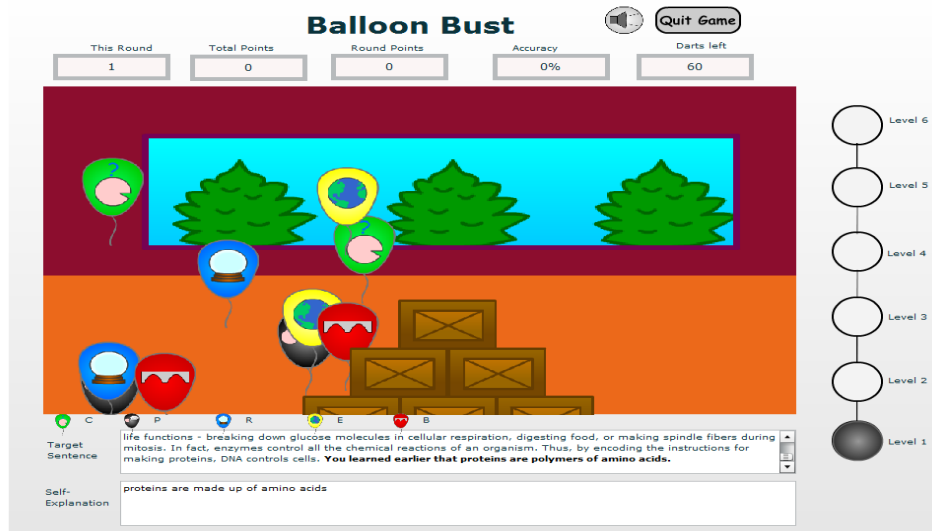


Figure 3. Screenshot of Balloon Bust

Current Study

In summary, self-regulation has been shown to be a critical skill for students' academic success. However, research in this area has typically focused on results emanating from self-report measures (Boekaerts & Corno, 2005; Cleary, 2006; Zimmerman, 1990, 2008). These static and often overt measurements of students' ability to regulate their behaviors often miss fine-grain behavior patterns that manifest while students engage within a learning task. Thus, dynamical analysis that focus on the fluid changes in nuanced behavior patterns are needed to gain a deeper understanding of how self-regulation emerges overtime and, ultimately, impacts learning outcomes. The current study uses two statistical techniques; random walks and Hurst exponents. The combination of these two techniques provides a novel means to visualizing and categorizing nuances in students' behavior patterns that emerge within system log-data across time. Such methodologies also afford us the opportunity to investigate how variations in dynamical patterns impact learning outcomes (i.e., self-explanation quality) within the context of iSTART-ME.

METHOD

Participants

The data that are analyzed in this paper were collected as part of a larger study that compared three conditions: iSTART-ME, iSTART-Regular, and a no-tutoring control (Jackson & McNamara, 2013). Participants in the current study are the subset of students from the original study who were assigned to the iSTART-ME condition. These participants included 40 high-school students from a mid-western urban environment. The students were, on average, 15.5 years of age, with a mean reported grade level of 10.4. Of the 40 students, 50% were female, and 17% were Caucasian, 73% were African-American, and 10% were other nationalities.

Procedure

The study comprised 11 sessions that included a pretest, 8 training sessions, a posttest, and a delayed retention test. During the first session, participants completed a pretest survey comprising a battery of measures, including an assessment of their prior self-explanation (SE) ability. During sessions 2 through 9, students engaged with the iSTART-ME system. Throughout these training sessions, students interacted with the full game-based menu, where they could choose to interact with generative practice games, identification mini-games, personalizable features, and achievement screens. Students completed a posttest during session 10 that included similar measures to the pretest. One week after completing the posttest, students returned to the lab for session 11, which consisted of a retention test that contained similar measures similar to the pretest and posttest (e.g., self-explanation ability).

Measures

Strategy performance. To assess self-explanation quality at pretest, posttest, and retention, students were asked to read through a text one sentence at a time and were then

prompted to provide a self-explanation for approximately 8 to 12 target sentences within each text. Students also generated self-explanations during training while interacting with the practice games in iSTART-ME. The quality of students' generated self-explanations was assessed through the use of a feedback algorithm that utilizes both latent semantic analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007) and word-based measures (McNamara, Boonthum, Levinstein, & Millis, 2007). This algorithm scores self-explanations on a scale ranging from 0 to 3. A score of "0" is assigned to poor self-explanations principally comprised of irrelevant information that is not contained in the text. A score of "1" is assigned to self-explanations related to the target sentence, but lacking elaborations using information from the text or prior knowledge (e.g., paraphrases). A score of "2" is assigned when self-explanations incorporate information from the text beyond the target sentence (e.g., include text-based inferences). Finally, a score of "3" suggests that a student's self-explanation incorporates information from both the text and prior knowledge. The assessment accuracy of this algorithm has been shown to be comparable to human ratings across a variety of texts (McNamara et al., 2007; Jackson, Guess, & McNamara, 2009).

System Interaction Choices

Students' interactions with iSTART-ME involved one of four types of game-based features, each representing a different type of game-based functionality within iSTART-ME:

1. *Generative practice games.* iSTART-ME includes three practice environments (Coached Practice, Map Conquest, and Showdown) that include generating self-explanations within the context of the games.
2. *Identification mini-games.* There are six identification mini-games that reinforce the learning strategies and goals presented to the student by asking the students to

identify the type of strategies used within self-explanations, but do not prompt the student to generate a self-explanation.

3. *Personalizable features.* Students have the opportunity to personalize features within the iSTART-ME environment. These customizable options include: editing an avatar, customizing the background theme, or changing their pedagogical agent.
4. *Achievement screens.* As students engage with the iSTART-ME system, they can earn points, win trophies, and advance to higher achievement levels. Within the main interface, students can view their progress in the system by scrolling over icons and opening achievement screens. When students choose to view any of these progress screens they are engaging with achievement screens.

Tracking the use of these four distinct features of iSTART-ME using log-data collected during the study affords the means to investigate patterns in students' choices across and within each type of interaction.

Quantitative Methods

To examine variations in students' behavior patterns within iSTART-ME, random walk analyses and Hurst exponents were calculated. Surrogate analyses were conducted to validate the interpretability of Hurst exponents. Hierarchical linear regressions were calculated to assess how students' behavior patterns influenced learning outcomes. The following section provides a description and explanation of random walk, Hurst, and surrogate analyses.

Random Walk Analyses

Random walk analyses were used in this study to visualize students' interaction patterns with iSTART-ME by examining the sequential order of students' interactions (i.e., choice of game-

based feature) with the four types of game-based features (i.e., generative practice games, identification mini-games, personalizable features, and achievement screens). Each of these feature types was assigned to an orthogonal vector on an XY scatter plot: generative practice games (-1,0), identification mini-games (0,1), personalizable features (1,0), and achievement screens (0,-1). Notably, these locations are random and are not associated with any qualitative value associated with the activity.

Each student’s walk was traced by first placing an imaginary particle at the origin (0,0). Every time that a student interacted with one of the four feature categories, the particle moved in a manner consistent with the vector assignment (see Table 1 for axis directional assignment). The use of these vectors allows us to define the movements that students make within the system. In the current study, vectors do not represent positive or negative dimensions; they simply provide a space on the grid to track users’ pattern of movements.

Table 1.

System interaction choice and corresponding axis direction assignment

System Interaction Choice	Axis Direction 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 illustrates what a random walk might look like for a student with five interactions. The starting point for all the walk sequences is (0,0); this is where the horizontal and vertical axes intersect (see # 0 in Figure 4). In this example, the first interaction that the

student engaged in was a mini-game; so, the particle moves one unit up along the Y-axis (see # 1 in Figure 4). The second interaction in which the student engaged was with a generative practice game, which moves the particle one unit left along the X-axis (see # 2 in Figure 4). The student's third interaction was with an achievement screen, which moves the particle one unit down along the Y-axis (see # 3 in Figure 4). The fourth hypothetical interaction was with another achievement screen, which again moves the particle one unit down along the Y-axis (see # 4 in Figure 4). Finally, for the fifth and final particle move, the student interacted with a personalizable feature, which moves the particle one unit to the right along the X-axis (see #5 in Figure 4). These simple rules were utilized for every interaction a student made within iSTART-ME. This analysis resulted in a custom walk for each of the 40 students.

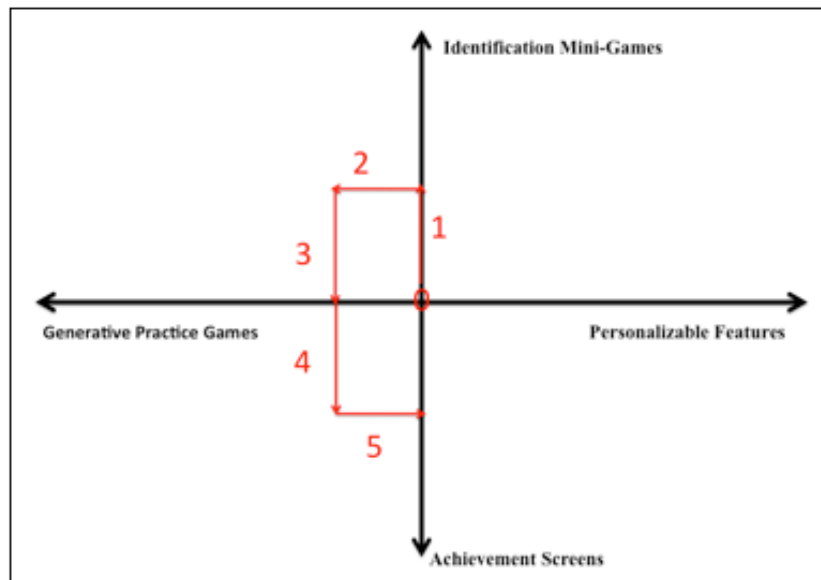
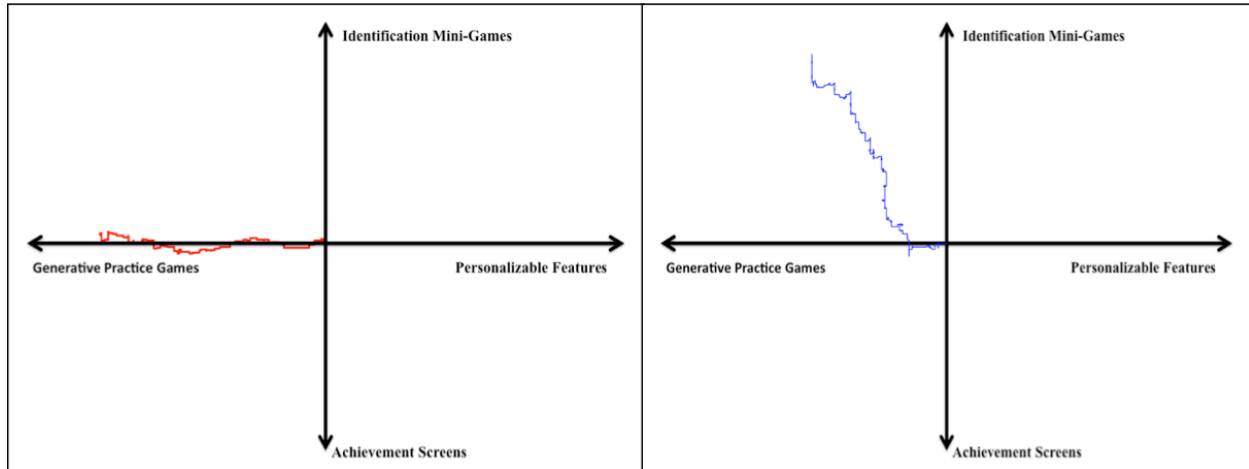


Figure 4. Example random walk with five interaction choices

Figures 5 and 6 illustrate two random walks that were generated using students' log-data. In Figure 5, the generated random walk reveals that this particular student interacted most frequently with both generative practice games and identification mini-games. This is demonstrated by the trajectory of their walk, as it hovers between the generative practice and

identification mini-games axes. Conversely, the student who generated the random walk in Figure 6 interacted most frequently with the generative practice games. This walk trajectory is primarily anchored along the generative practice axis. These two contrasting figures demonstrate how the log-data can be used to generate a spatial representation of each student’s time in the system.



Figures 5 & 6. Random walks generated from two different students’ log-data

Using the data resulting from the random walk analyses, distance times series were constructed for each student by calculating a measure of Euclidean distance for each step in the walk. Distance was calculated from the origin to each step with equation (1) where y represents the particles place on the y-axis, x represents the particles place on the x-axis, and i represents the i th step in the walk:

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

Hurst Exponents

To classify the tendency of students’ interaction patterns based on the distance time series analyses, Hurst exponents were calculated using Detrended Fluctuation Analysis (DFA; Peng et

al., 1994). DFA estimates Hurst by integrating a normalized time series and then dividing the series into equal intervals of length, n . Each interval is fit with a least squares line and then the integrated series is *detrended* by subtracting local predicted (i.e., those for each interval) values from the integrated time series. The process is repeated for intervals of different lengths, increasing exponentially (e.g., by powers of 2). For each interval size, the characteristic fluctuation, $F(n)$, is calculated as the root mean square deviation of the integrated time series from local least squares lines. $\log_2 F(n)$ is then regressed onto $\log_2(n)$; the slope of the regression line gives the Hurst exponent, H . The interpretive index for Hurst is as follows: $0.5 < H \leq 1$ indicates persistent (deterministic or controlled) behavior, $H = 0.5$ signifies random (independent) behavior, and $0 \leq H < 0.5$ denotes antipersistent (corrective) behavior.

Surrogate Analysis

A surrogate analysis was calculated to validate the Hurst exponents generated by the DFA (Theiler, Eubank, Longtin, Galdrikian, & Farmer, 1992). Surrogate analysis tests the null hypothesis that the observed scaling behavior is an artifact of a truncated observational window. That is, the analyzed time series may actually be a random process that merely appears to exhibit persistent- or antipersistent-like behavior over a short interval. If so, randomly shuffling the series should not affect the scaling structure. Conversely, if the scaling behavior is genuine, then shuffling the time series should deteriorate the scaling structure. In surrogate analysis, each analyzed time series is shuffled many times (in our case 40) and DFA is performed on each of the shuffled series. Summary statistics are then calculated for the results of the shuffled time series analyses and those are compared with the intact series analyses using a t-test. A significant difference allows for the rejection of the null hypothesis and provides support for the interpretation of the scaling behavior.

RESULTS

Hurst Exponents and Surrogate Analyses

To characterize how students interacted with the system, Hurst exponents were calculated using DFA and students' distance time series derived from the individual random walks. Hurst exponents suggested that students varied considerably from weakly to strongly persistent (range =0.57 to 1.00, $M=0.77$, $SD=0.11$). A surrogate analysis was then conducted to assess the reliability of the Hurst exponents. The surrogate analysis revealed that Hurst exponents derived from intact series differed from those calculated on shuffled time series, $t(43)=150.63$, $p<.001$, suggesting that the Hurst exponents characterizing students' interaction patterns were reliable.

Interaction Choices

To examine the relation between Hurst exponents (i.e., as a measure of deterministic or random tendencies) and students' frequency of interaction choices we conducted Pearson correlations. Results from this analysis revealed that students' Hurst exponents were not significantly related to students frequency of interactions with generative practice games ($r=.25$, $p=.11$), identification mini-games ($r=-.12$, $p=.45$), personalizable features ($r=-.06$, $p=.70$), or achievement screen views ($r=-.24$, $p=.13$). These results indicate that students' interaction patterns within the system were not related to any specific feature.

Learning Outcomes

Using Pearson correlations, we measured the strength of the relation between the Hurst exponents (i.e., as a measure of deterministic or random tendencies) and students' self-explanation scores during training, as well as at posttest and retention test scores (see Table 2). Results from this analysis revealed that students' Hurst exponents were significantly related to self-explanation scores during training ($r=.51$, $p<.001$) and at retention ($r=.31$, $p=.05$). However,

there was no relation between students' Hurst exponents and their self-explanation scores at posttest ($r=.09, p=.59$). These results indicate that when students interacted in a more controlled way within the system, they generated higher quality self-explanations during training and at retention.

Table 2.

Correlations between self-explanation scores and Hurst exponents

Strategy Performance	<i>r</i>
Training self-explanation score	.51**
Posttest self-explanation score	.09
Retention self-explanation score	.31*

p = .05* , p < .001**

To further investigate how interaction patterns impacted daily learning outcomes (i.e., training self-explanation scores), we used a hierarchal linear regression model to factor out students' pretest self-explanation score. In model one of this analysis, we used pretest self-explanation scores to predict daily training self-explanation scores. Results from this analysis revealed that pretest self-explanation score was a significant predictor of students' daily self-explanation scores ($R^2=.26, F(1,38)=13.60, p<.01$; see Table 3). In model two, we examined the degree to which students' Hurst exponents predicted daily self-explanation scores over and above pretest self-explanation score. Results from this analysis indicated that Hurst exponents were a significant predictor of daily self-explanation scores over and above pretest self-explanation score ($R^2=.44, F(1,37)=11.93, p<.01$; see Table 3). This analysis demonstrated that students' Hurst exponents accounted for 18% of the additional variance in students' daily self-explanation quality over and above pretest self-explanation score.

Table 3.

Hierarchical linear regression analyses predicting daily self-explanation quality

Variable	B	SE	β	ΔR²
Model 1				.26**
Pretest self-explanation	.41	.11	.51**	
Model 2				.18**
Pretest self-explanation	.35	.10	.43**	
Hurst Exponent	1.89	.55	.43**	

* $p < .05$; ** $p < .01$

A similar hierarchical linear regression model was conducted to investigate the degree to which interaction patterns impacted performance at the retention test over and above students' pretest self-explanation score. In model one of this analysis, we used pretest self-explanation scores to predict retention self-explanation quality. Results from this initial analysis demonstrated that pretest self-explanation quality was a significant predictor of students' retention self-explanation quality ($R^2=.22$, $F(1,38)=10.67$, $p<.01$; see Table 4). In model two, we examined the degree to which students' Hurst exponents predicted their retention self-explanation quality over and above pretest self-explanation quality. Results from this analysis indicated that students' Hurst exponent did not significantly predict the quality of their retention self-explanation quality over and above pretest self-explanation quality ($R^2=.27$, $F(1,37)=2.70$, $p=.10$; see Table 4).

Table 4.

Hierarchical linear regression analyses predicting retention self-explanation quality outcomes

Variable	B	SE	β	ΔR^2
Model 1				.22**
Pretest self-explanation	.47	.14	.47**	
Model 2				.05
Pretest self-explanation	.43	.61	.43**	
Hurst Exponent	1.28	.77	.23	

* $p < .05$; ** $p < .01$

DISCUSSION

A predominant assumption by many researchers and educators is that students' ability to regulate their behaviors during learning has a positive and important impact on their academic success (Hadwin et al., 2007; Sabourin et al., 2012; Zimmerman, 1990). However, investigations of these relations tend to rely heavily upon self-report measures. A common concern about this methodology is that self-report measures do not adequately capture the fine-grained changes that occur in students' behaviors over time. Hence, nuanced and dynamical measures are needed to gain a deeper understanding of students' ability to control, monitor, and regulate their behaviors (Hadwin et al., 2007). Log-data has been previously used along with self-report measures to analyze students' self-regulatory behaviors at more fine-grained levels (Hadwin et al., 2007; Sabourin et al., 2012). The work presented here builds upon these findings by conducting dynamical analyses of system log-data to investigate the extent to which students' behaviors

exhibit controlled properties. These initial analyses explore how dynamical techniques can potentially act as a form of stealth assessment within systems, such as iSTART-ME. Such assessments have a strong potential to deepen our understanding of the relations between learning outcomes and sequences in students' behaviors within adaptive environments.

Researchers have previously argued that one aspect of SRL is the ability to control one's own behaviors and to not act in an impulsive manner while engaging in learning tasks (McCombs, 1989; Schunk, 2008; Schunk & Zimmerman, 2003); thus, patterns of choice within systems can be interpreted as revealing aspects of self-regulation. The current study made use of novel methodologies by employing random walk and Hurst exponent analyses in an attempt to capture each student's interaction pattern within iSTART-ME. Past research using Hurst exponents points to the use of this scaling variable as an indicator of the degree to which students' behaviors are controlled and deterministic (Mandelbrot, 1982; Van Orden, Holden, & Turvey, 2003). Specifically, when students act in a deterministic manner, they are exhibiting persistent and controlled behavior patterns. Conversely, when students engage in random behaviors, they are not interacting with purpose, control, or persistence. These tendencies across long periods of time reveal trends in how students approach learning tasks. Therefore, this work begins to shed light upon the dynamical nature of SRL behaviors that students exhibit while interacting with adaptive systems.

Results from the current study fall in line with previous work that has shown that students' ability to regulate their learning behaviors has a positive impact on learning outcomes (Butler, & Winne, 1995; Pintrich, & De Groot, 1990; Zimmerman & Schunk, 1989; Zimmerman, 1990). Specifically, we found a significant positive relation between controlled patterns of interactions (i.e., Hurst scores) and self-explanation quality assessed during training and at the

retention test (though not with performance at posttest). The relation between Hurst scores and daily training performance held when also statistically accounting for students' prior self-explanation ability. However, this was not the case for performance during the retention test. Thus, students' ability to control and regulate their behaviors seems to be important for immediate learning outcomes, but does not impact long-term learning. These results suggest that when students are given more control over their environment there are some potential consequences, at least for immediate performance. This may be especially important within game-based environments, as they often offer students numerous opportunities to control their trajectory within the system (Sabourin et al., 2012; Snow et al., 2013). Such agency may or may not be appropriate depending on the learning goal embedded within the environment.

This study serves as a starting point for scientists to apply dynamical techniques to system log-data as a way to trace and classify students' interactions. These analyses are intended as a seed for future studies by providing evidence that dynamical methodologies show strong promise in providing online stealth indicators of self-regulation. The next steps in this research agenda necessarily include further confirmatory studies demonstrating concurrent validity. For example, an obvious extension of the current work will be to include self-report measures that have been traditionally used to assess self-regulation. The outcomes of such assessments can then be compared to those provided by dynamical assessments. Notably, however, the results of such studies may be inconclusive given that the one supposition of the current research is that self-report measures are fundamentally flawed. Thus, the horizon of future research may lie in establishing the respective utility of using static versus dynamical assessments of self-regulation during learning.

Another future direction of study regards the practical use of this approach. Ultimately, the purpose of using dynamical measures is to capture self-regulation in real time. Hence, the true test lies in the implementation of these measures within adaptive learning environments to evaluate their utility in those contexts. For example, one crucial question for future research regards the use of visualization and dynamical techniques as a means to unobtrusively assess self-regulated behavior patterns. Such analyses will be especially valuable if systems are able to recognize non-optimal patterns and steer students toward more effective behaviors. For instance, if a student is engaging in a random interaction loop, it may be beneficial for adaptive learning environments to have the capability to recognize these patterns and prompt the student toward a more regulated trajectory.

In conclusion, this study explored the use of two dynamical methodologies to unobtrusively assess self-regulated behavior and its impact on learning within a game-based environment. These analyses are among the first attempts to examine nuances in students' log-data to capture tendencies in online behaviors and subsequent interaction patterns across time. Student models rely on understanding the relation between students' abilities and performance. We expect tracking and modeling interaction trends over time to be crucial to improving adaptivity within systems that provide students with agency over the environment. Overall, these findings afford researchers the opportunity to understand the dynamical nature of SRL and its impact on student learning.

Chapter 4

GENERAL DISCUSSION

Students' ability to exert agency over their learning experiences by setting goals and strategic plans has been linked academic success (Hadwin & Winne, 2007; Zimmerman, 2008). However, students vary in their ability to effectively control and regulate their behaviors (Zimmerman, 1990). Traditionally, self-report measures have been used to measure fluctuations in this ability. However, these static (i.e., unchanging) measures are unable to assess potential variability and fine-grained nature of these behaviors. One way to capture variance in students' ability to control their behaviors and act in a strategic way is through the analysis of in-system log-data (Sabourin et al., 2012). Specifically, log-data from game-based systems has strong potential to provide researchers with an opportunity to covertly assess SRL behaviors (Sabourin et al., 2012; Snow 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. Consequently, these environments provide researchers with an opportunity to examine the impact that students' ability to exert control over their learning process has on learning outcomes.

The work presented here utilized log-data as a means to capture both course-grained and fine-grained patterns of strategic planning reflected in students' choices within the iSTART-ME interface. In Study 1, students' proportional use of their in-game currency was investigated. The proportion of system currency that students used was termed *spendency*. This measure reflects the degree to which students spent their in-system currency (iBucks). Spendency was designed to calculate a trait that students demonstrate within the game-based system. As such, this measure does not reflect the number of iBucks earned, instead it focuses on students' propensity to use their earned iBucks. Thus, if a student exhibited high spendency, they were spending their earned

iBucks at a high rate. Such students were essentially spending iBucks as soon as they earned them. In the current study, it was hypothesized that this behavior was indicative of a lack of strategic planning.

Within the SRL literature, strategic planning is often associated with controlled behaviors that are aimed at achieving learning goals (Zimmerman, 1990). If students are spending a high proportion of their earned iBucks they seem to be placing a higher importance on the game-based features rather than on the practice and acquisition of the target skills. This priority is reflected in the interaction loop presented in Figure 6 (page 25), where students engage with a generative practice game and then immediately spend their earned iBucks on personalizable features. In the current study, it was hypothesized that a student with high spendency did not display a strategic plan of action, but instead game-based features seduced them. Indeed, students who successfully self-regulate would be monitoring their interactions within the iSTART-ME system and would be more likely to identify that the personalizable features are seductive and tangential to the actual learning task.

This hypothesis is further supported by the performance and learning outcomes findings from Study 1. Specifically, students who had a higher spendency generated lower quality self-explanations during training and performed worse on a transfer task. Thus, students who did not regulate their interactions with the seductive distracters performed worse than those students who did set a plan of action and controlled how often they spent their earned iBucks. An alternative hypothesis could be that students spent a higher proportion of iBucks because they were bored or unmotivated. However, further analyses presented in Study 1 revealed no significant relation between students' spendency and their self-reported enjoyment or motivation. These results

suggest that students' propensity to spend their earned iBucks does not seem to be the result of boredom or disinterest.

In this thesis project, Study 1 examined students' behaviors in a course-grained manner. However, self-regulation is argued to be dynamical in nature, and hence, evolving overtime (Zhou, 2013). The changing nature of SRL is one of the reasons that researchers have struggled to measure this construct. Study 2 attempts to capture fine-grained variations in students' behaviors through the use of dynamical systems theory analysis. Specifically, Study 2 made use of three novel methodologies by employing random walks, Euclidean distances, and Hurst exponents in an attempt to capture fluid patterns that emerged in students' behaviors within the iSTART-ME interface. Random walk analyses provided a visualization of each user's interaction trajectory. Euclidian distance measures provided a quantification of changes in users' choices across time. Finally, the Hurst exponent classified the tendency of these interaction trajectories. Past research using Hurst exponents has revealed that this measure is an indicator of the degree to which time series fluctuations are random (Hurst, 1951). The current analysis is one of the first studies to use Hurst as a means to provide a stealth assessment of students' patterns of interactions within a tutoring environment.

Within Study 2, when students produced a Hurst exponent close to 1, it was interpreted as the student exerting a controlled and persistent pattern of interactions. These students are hypothesized to be setting a plan of action for interacting in the system. To illustrate these behaviors, transitional probability analysis similar to the one presented in Figure 6 for Study 1, were conducted. This analysis examined how one student who had a Hurst exponent of .98, interacted with various features in the iSTART-ME system. This relation is depicted in Figure 1. These results revealed that almost 60% of this student's interactions were with the generative

practice games. There was little tendency to interact with other features and when engaging with another game-based feature, the student transitioned back to the generative practice games afterwards. Thus, the student seemed to be acting in a decisive manner, consistently interacting with generative practice games or transitioning back to generative practice games after engaging with another feature.

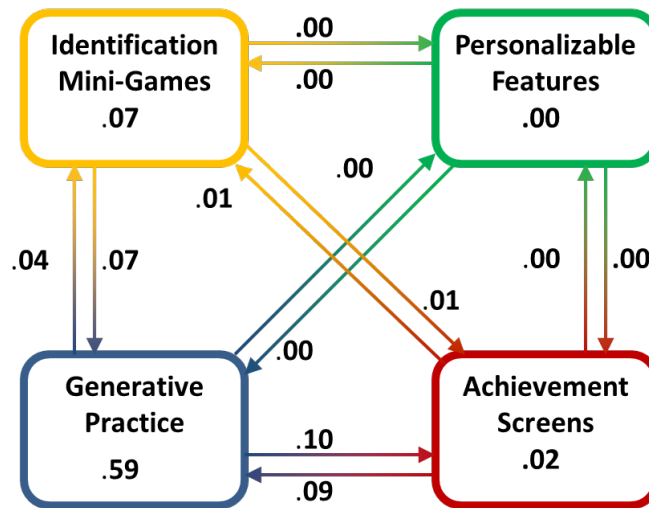


Figure 1. Transitional Probability of High Hurst Student

Conversely, in Study 2 when students produced a Hurst exponent closer to .5, it was interpreted as the student acting in an impetuous or random manner. These students did not exert a strategic way of interacting with the system. They may have jumped around from feature to feature unaware or unsure of what to do next. In Study 2, it was hypothesized that students with Hurst exponent scores closer to .5 were not acting in a decisive or strategic manner. To illustrate what these behaviors may look like, a transitional probability analysis was conducted for one student who had a Hurst exponent of .60 (Figure 2). This analysis revealed that the student with a low Hurst score explored more of the system interface than the student with a Hurst score of .98. However, the interaction pattern was more spread out and less predictable compared to the student depicted above. It is hypothesized that the student in Figure 2, was not acting in a

decisive manner and as such, may not have been setting a strategic plan of action within the iSTART-ME system. This hypothesis is supported by the target skill performance findings from Study 2. Specifically, students who displayed a more decisive interaction trajectory within iSTART-ME generated higher quality self-explanations during training than students who jumped around the system more frequently.

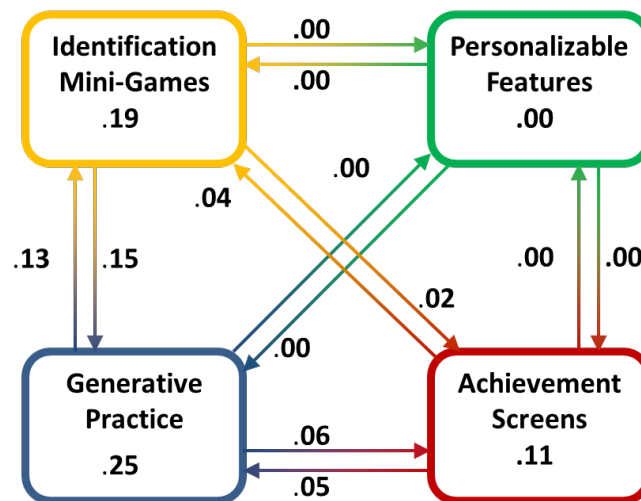


Figure 2. Transitional Probability of Low Hurst Student

Findings from both Study 1 and Study 2 demonstrate that log-data from game-based environments can reveal students’ behavioral tendencies as they emerge across time. Traditionally, self-regulation research has relied on static self-report measures. These measures may not capture the fine-grained and dynamical behaviors that evolve throughout the learning process (Hadwin & Winne, 2007). This thesis adds to the literature by demonstrating that fine-grained behavior patterns are related to target skill acquisition and long-term learning. Without relying on self-report measures or students’ perception of events, these two studies presented here capture variation in behaviors that emerge organically throughout the students’ time within iSTART-ME. Self-regulation is a dynamic process that changes and evolves as students become more engaged in the learning process. The work presented here reveals that log-data affords

researchers a means to capture these emergent behaviors and the impact that they have on learned skills.

Future Directions and Follow-Up Studies

The methodologies and analyses presented in this thesis serve as a starting point for covertly capturing one aspect of self-regulation (i.e., strategic planning). The use of Hurst analyses within Study 2 provide a means to quantify students' interaction trajectories within iSTART-2. Hurst exponents were chosen as the primary methodology within Study 2 because they provide a long-term correlation of fluctuations in students' choices within the system. However, this analysis does not take into account the directionality of each student's random walk pattern. In the future, it will be important to explore other time series analysis techniques that can capture directionality of each students walk, such as recurrence and derivative slope analyses. Future work will also focus on the amount of time that students spend on each feature.

Although the analyses presented here provide further evidence that fine-grained patterns within log-data can shed light upon various behavioral trends that may manifest during the learning process, confirmatory studies are still needed. In the future, follow-up work will be conducted to demonstrate the relation between traditional measures of control and regulation (i.e., self-reports) and these novel statistical techniques. Such work will potentially establish the respective utility of log-data analysis and traditional measures of learning behaviors (as well as the subsequent intent behind those behaviors). Two future studies have been planned to further establish the validity of log-data pattern analysis as a measure of SRL. These studies are described in the following sections.

The ultimate goal of this research line is to utilize log-data in real-time as a form of stealth assessment. With the information provided by these stealth assessments, ITSs may be

able to recognize non-optimal learning paths and subsequently steer students toward more optimal learning behaviors. This transition into more optimal learning patterns may also serve as a means to improve students' metacognitive awareness. When students are more aware of the task's learning goal and how their behaviors are influencing that goal, they are said to have metacognitive awareness (Muraven, Baumeister, & Tice, 1999). This trait has been positively linked to SRL (Muraven, Baumeister, & Tice, 1999). Thus, the real-time analysis of log-data may prove to be a useful way to increase students' metacognitive awareness. Specifically, real-time log-data analysis may be useful in providing students with individualized examples of on their optimal and non-optimal learning behaviors. A third follow-up study has been planned to test this assumption by examining how log-data can be used to help students recognize non-optimal patterns and transition them toward more effective behaviors.

Follow-up Study 1. The first follow-up study will be a correlational study aimed at replicating the findings presented in Study 2 of this thesis. This study will also include traditional measures of SRL as a means to establish the link between the dynamical analysis of log-data and previously validated measures of SRL.

Procedure follow-up Study 1. Students in this study will complete one 3-hour session consisting of a pretest (approximately 20 minutes), strategy training (approximately 20 minutes), game-based practice within iSTART-2 (new iteration of iSTART-ME; approximately 2 hours), and a posttest (approximately 20 minutes). During the pretest, students will answer two questionnaires that assess their prior strategy and self-regulative ability. At pretest, students will also be asked basic demographic information (e.g., sex and age). During training, students will watch the iSTART training videos that instruct them on self-explanation strategies and their applications. After training, students will be transitioned into the game-based practice portion of

the experiment. In this section, students will be allowed to interact freely within the iSTART-2 interface. Finally, at posttest students will complete a self-explanation task similar to the one presented in the pretest.

Participants follow-up Study 1. For this correlational design, a power analysis using a desired .80 power indicated that a sample of $n = 82$ participants is needed to detect a medium effect size. Participants will be undergraduate students from the Psych 101 subject pool.

Measures.

Self-explanation quality. Students' self-explanation ability will be assessed at both pretest and posttest using two science texts, which will be counterbalanced among students. These two texts are comparable in grade level. Each student will be shown eight target sentences and after each they will be prompted to self-explain the presented text. During training, students' self-explanation ability will be assessed through the generative practice games. All self-explanations will be scored using the previously mentioned iSTART algorithm.

Self-regulation ability. Students' self-regulation ability will be assessed using the attention control self-report scale (ACS, Derryberry & Reed, 2002). This scale is a 20-question likert style survey that asks students a series of questions designed to assess their own intuition about their ability to control and direct their attention.

System interaction choices. Within iSTART-2, students can chose to interact with a variety of features that fall into one of four types of game-based feature categories (i.e., generative practice, identification mini-games, personalizable features, and achievement screens). Every interaction choice a student makes while they engage within the system interface will be sequentially logged.

Quantitative methods. In an attempt to replicate the findings of Study 2 of this thesis, random walks, Euclidean distances and Entropy analyses will be conducted to visualize and classify variations in students' choice patterns. After classifying students' patterns of interactions, these relations will be measured between these variations and self-explanation quality during training and at posttest.

Random walks and Euclidian distances. Similar to Study 2 of the thesis, each student's log data will be used to generate a random walk that provides a visual illustration of their time within the system. Distance measures will be used to mathematically quantify fluctuations within these walks. The random walk and Euclidian distance measure will be calculated using the same procedure explained in Study 2 of the thesis.

Entropy analysis. Although Hurst exponents were used in Study 2 to classify the fluctuations in students' walks, this type of analysis will not be possible for a one-session study. The Hurst exponent analyses requires multiple data points and for a one-session study it is hypothesized that students will not have a sufficient number of interaction points to calculate a reliable Hurst exponent. To address this issue, the current study will use an Entropy analysis to classify students' patterns of interactions. Entropy analyses and Hurst exponents have both been previously used to measure random, controlled, and ordered processes (Grossman, 1953; Hurst, 1951). Within the proposed study, Entropy will afford the opportunity to gain a deeper understanding of how students' choice patterns reflect controlled and ordered processes across fewer interaction data points than required for the Hurst. In the proposed study, when students produce a low Entropy score they will be interpreted as exhibiting persistent and controlled behavior patterns. Conversely, a high Entropy scoring student will be interpreted as having a

random interaction pattern within the system that displays no discernable purpose or element of control.

Proposed analyses. To examine the relation between ordered patterns of interactions (i.e., Entropy) and a traditional measure of SRL (i.e., ACS), a correlational analysis will be conducted. Similar correlational analyses will be used to examine the relation between Entropy and self-explanation quality during training and at posttest.

Expected results. It is hypothesized that the follow-up Study 1 results will reveal a significant relation between Entropy scores and self-explanation quality during training. However, it is not expected that Entropy will be related to posttest performance. These expected results mirror the findings presented in Study 2 of this thesis. It is also hypothesized that there will be a significant relation between students' self-reported ability to control their attention and Entropy scores. These results would begin to validate the hypothesis that ordered patterns of interaction are related to previously validated measures of SRL.

Follow-up Study 2. The first follow-up study was a correlational analysis aimed at replicating findings from Study 2 of the thesis. The second follow-up study will further examine the influence of ordered behavior patterns (i.e., Entropy scores) on target skill acquisition (i.e., self-explanation quality) as a function of agency. A main component of self-regulation is students' ability to take agency over their learning path and set their own strategic plan (Zimmerman, 2008). The second follow-up study will be designed to test that assumption by removing students' agency by randomly assigning them to an iSTART-2 interaction pattern from a student who participated in follow-up Study 1. Hence, students in Study 2 will be assigned to trajectory that was classified as either ordered or random trajectory in Study 1, resulting in

between-subjects manipulation of assigned trajectory. The design and experimental manipulation are further described below.

Design. Follow-up Study 2 includes a between-subjects variable (i.e., assigned trajectory) with two conditions that are designed to examine the effects of interaction pattern on target skill acquisition. These two conditions will be created by yoking all of the interaction paths previously generated by participants in the Study 1 follow-up. These yoked interaction patterns will have all been previously assigned an Entropy score from the follow-up Study 1 analysis. A median split will be conducted on these scores to create two groups of interaction patterns (i.e., random pattern condition and controlled pattern condition). In the follow-up Study 2, all students will be randomly assigned to one of these two experimental conditions. After being assigned to a condition, students will be given a list of interactions (i.e., interaction pattern) that they must follow in sequential order while engaged within the iSTART-2 environment.

Procedure follow-up Study 2. The procedures for the second follow up study are identical to the follow-up Study 1 procedures for pretest, strategy training, and posttest. The only difference occurs when students are transitioned into the game-based practice portion of the experiment. In this section, instead of having free choice as did the students in follow-up Study 1, students in follow-up Study 2 will be randomly assigned to an interaction trajectory from a student in Study 1.

Participants follow-up Study 2. For this study, 82 participants will be recruited. This ensures that each previously generated interaction pattern from follow-up Study 1 will be yoked to a student in follow-up Study 2. All participants will be undergraduate students from the Psych 101 subject pool.

Measures. All measures for follow-up Study 2 will be identical to the measures presented within follow-up Study 1.

Proposed analyses. To examine the impact of ordered patterns on self-explanation quality, two one-way ANOVAs will be conducted. The independent variable will be students' assigned trajectory (i.e., random vs. controlled) and the dependent variables will be self-explanation quality at training and at posttest, respectively.

Expected results. It is hypothesized that the results from the Study 2 follow-up will reveal no significant differences in self-explanation quality at training or posttest between the two conditions (i.e., random and controlled interaction patterns). This null result is predicted because it is hypothesized that ordered interaction patterns only influence learning outcomes when students' exert agency and create their own trajectory. Thus, assigning students a statistically ordered path without any context may seem just as random to the student as a statically random path.

Proposed meta-analysis for follow-up Studies 1 and 2. A 2 x 2 ANOVA will be conducted to examine variations in target skill performance between the two follow-up studies. This comparison will be conducted to examine variations in self-explanation quality during training as a function of agency. In this analysis, average self-explanation quality during training will be the dependent variable. The two independent variables will be study (Study 1 or Study 2) and condition (ordered interactions or random interactions). Students who participated in follow-up Study 1, will be assigned to a condition (controlled or random) post-hoc based on the median split conducted prior to follow-up Study 2.

Meta-analysis expected results. It is hypothesized that there will be a main effect of study. Thus, students in follow-up Study 1 will generate higher quality self-explanations during

training compared to students in follow-up Study 2. A main effect of condition is not expected, as it is hypothesized that ordered interaction patterns will only influence self-explanation quality at training when the students have generated their own interaction trajectory. Finally, it is hypothesized that there will be a significant interaction between study and condition. Thus, ordered interaction patterns will only influence self-explanation quality at training when students' have generated their own interaction path. These hypotheses are supported by SRL literature that reveals the importance of personal agency when creating a plan of action for learning tasks (Zimmerman, 2008). Strategic planning is based on personal agency and students' assessment of what they need to do to accomplish a learning goal. The results from these studies will begin to reveal that personal agency is an important component of the relation between interaction patterns and target skill acquisition.

Follow-up Studies 1 and 2 general discussion. The overall goals of the first two proposed follow-up studies are to establish validity between the dynamical analyses of log-data and to begin to tease apart the cognitive processes that may be underlining these fluid interaction patterns. In particular, these studies are designed to investigate how students' agency in setting a plan of action influences the relation between controlled patterns of interaction and target skill acquisition.

System design implications. The ultimate purpose of using log-data is to capture students' behavioral trends that occur in real time and relate those trends to learning outcomes. Thus, the true measure of the applicability of these assessments 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 guiding students toward more ordered behavior trajectories.

Follow-up Study 3. Studies 1 and 2 investigate the relation between traditional measures of control and regulation (i.e., self-reports) and log-data analysis. These studies also examine how students' ability to exert agency over their interaction trajectory impacts learning outcomes. This work will 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 stealth assessments in real time and guide students toward optimal learning behaviors.

SRL research has shown that students' ability to control and regulate their behaviors can evolve over time (Glaser, & Brunstein, 2007). One factor shown to impact the emergence of this skill is metacognitive awareness (Muraven, Baumeister, & Tice, 1999). When students are aware of how to approach a learning task, and the subsequent strengths and weakness associated with that behavior, they perform better than students who are unaware. The real-time analysis of log-data may be useful in providing students with individualized examples of on their optimal and non-optimal learning behaviors, thus, improving their metacognitive awareness. Follow-up Study 3 will be designed to test the utility of online-measures as a way to increase metacognitive awareness and long-term learning outcomes.

Design. Follow-up Study 3 is a 2 (mastery) x 2 (Entropy) factorial design (see Figure 3). This study is designed to examine how the utilization of stealth measures (i.e., mastery level and Entropy score) can be used in real time to transition students toward more optimal learning paths and the ultimate impact of prompting on long-term learning outcomes (i.e., self-explanation and

reading comprehension skill). This study will use two online measures (mastery and Entropy) as stealth measurements of students' skill performance and controlled behavior patterns. In this study, mastery captures students' demonstrated knowledge of the target skill (self-explanation) and Entropy captures variations in students' ability to control their interaction patterns.

<u>Mastery Threshold</u>	<u>Entropy Threshold</u>	
	None	1.5
None	No Transitions	Transition Based on Entropy
2.0	Transition Based on Mastery	Transition Based on Entropy and Mastery

Figure 3. Factorial design for follow-up Study 3

Mastery. Students' mastery will be calculated using their daily average self-explanation scores. In the current study, the threshold for mastery is an average self-explanation score at or above 2.0. This score was chosen for mastery because a 2.0 is indicative of the use of information beyond the text (i.e., use of prior knowledge; Jackson & McNamara, 2012).

Entropy. Students' Entropy score will be calculated in real-time by using the same quantitative methodology proposed in follow-up Study 1. Within this study, Entropy will be calculated after each interaction a student makes within the system starting with their first choice. This will result in a *running* Entropy score that is updated after every choice a student makes. An Entropy threshold will be set at 1.5, this number was chosen because it is the hypothesized median split value from follow-up Study 1, which differentiated the random and controlled interaction conditions. These two online measures will be used in real time to transition students (when necessary) to change their behaviors within the iSTART-2 system. These transitions will occur starting in session 3 (day 2 in iSTART-2) and they will last throughout students' time within the system.

Conditions. Within the current study, the independent variables of mastery and Entropy each include two levels, thus, resulting in four between-subject conditions (see Figure 3). In the no transition condition (top left of Figure 3), students will not receive any behavioral modification. Thus, there is not a threshold of Entropy or mastery for these students. In the transition based on Entropy condition (top right of Figure 3), students will be steered (i.e., put into) to the coached practice feature, if their running Entropy score is below 1.5. In the transition based on mastery condition (bottom left of Figure 3), students will be steered to the coached practice feature if their average self-explanation score falls below 2.0. Finally, in the transition based on Entropy and mastery condition (bottom right of Figure 3), students will be steered to the coached practice feature, if they have an Entropy score lower than 1.5 and if their average self-explanation score falls below 2.0. In the transition based on Entropy and mastery condition, both thresholds must be met (mastery and Entropy threshold) for a student to be steered into coached practice. The coached practice feature was chosen because it has no game-features and provides students with feedback as they generate self-explanations. Before students are transitioned to coached practice they will receive a pop-up message that is condition specific (see Table 1).

Table 1.

Follow-up study 3 pop-up messages.

Condition	Threshold	Message
No Transition	N/A	N/A
Transition Based on Entropy	1.5 Entropy	You seem to be jumping around the system frequently. You will now be transitioned to coached practice to apply the self-explanation training you have received.
Transition Based on Mastery	2.0 mastery	You are generating low quality self-explanations. You will now be transitioned to coached practice to apply the self-explanation training you have received.
Transition Based on Mastery and Entropy	1.5 Entropy 2.0 mastery	You seem to be jumping around the system frequently and you are generating low quality self-explanations. You will now be transitioned to coached practice to apply the self-explanation training you have received.

Participants. For this 2x2 factorial design, a power analysis using a desired .80 power indicated that a sample of $n = 128$ participants is needed to detect a medium effect size (i.e., 32 students per condition). Participants will be undergraduate students from the Psych 101 subject pool.

Procedure. This study will be an 8-session experiment, which consists of a pretest, five training sessions, a posttest, and a retention test. During the first session, participants will complete a pretest survey that includes demographics and measures of prior self-explanation ability and reading ability. During the five training sessions, participants will engage within the iSTART-2 interface for approximately 1 hour per session. After the five training sessions, students will complete a posttest that includes similar measures to the pretest. Finally, in the final session, students will complete a retention test that includes a measure of reading comprehension ability.

Prior to the training portion of the experiment all students will be randomly assigned to one of the four aforementioned experimental conditions. This condition assignment will determine how “free” a student is to interact with the system and what Entropy or mastery thresholds will be used to direct their interactions within the system (when necessary).

System interface set-up. For follow-up Study 3, the iSTART-2 system interface will start off locked, where students can only engage with the generative practice games. Then as students earn points within the system the other types of game-based features will become unlocked. This interface design is similar to the design used in Study 1 and Study 2 of the thesis. This design ensures that students do engage with generative practice games and do not spend their entire time within the system on an off-task feature (i.e., personalizable features).

Measures.

Self-explanation quality. Students’ self-explanation ability will be assessed at both pretest and posttest using two science texts, which will be counterbalanced among students. These two texts are comparable in terms of grade level. Each student will be shown eight target sentences and after each they will be prompted to self-explain the presented text. During training, students’ self-explanation ability will be assessed through the generative practice games. All self-explanations will be scored using the previously mentioned iSTART algorithm.

Prior reading ability. Students’ reading comprehension ability will be assessed at pretest using the Gates-MacGinitie Reading Comprehension Test (MacGinitie & MacGinitie, 1989). This test includes 48 questions designed to assess the general reading ability of each student by asking them to read a passage and then answer comprehension questions about the passage.

Reading comprehension. Students’ reading comprehension will be measured through an open-ended comprehension measure. In this comprehension measure each student will be asked

to read a science text. After they finished reading, students will then be presented with a series of open-ended questions that will be based on their recollection of the text. These questions will assess both low-level and deep-level text comprehension.

Proposed analyses. For the initial analyses within this study, 2 x 2 ANOVAs will be conducted to examine how the experimental manipulations (i.e., transitions based on system thresholds) influenced students' performance on self-explanation quality at posttest and reading comprehension skill during retention. In the first analysis, the independent variable will be mastery and Entropy thresholds, whereas the dependent variable will be posttest self-explanation quality. This analysis will examine the main effects of mastery and Entropy along with the interaction of these two variables on posttest self-explanation quality.

A second 2 x 2 ANOVA will be utilized to examine the influence of mastery and Entropy thresholds on reading comprehension skill at retention. In the second analysis, the independent variables will be mastery and Entropy thresholds, whereas the dependent variable will be reading comprehension skill at retention. This analysis will examine the main effects of mastery and Entropy and with the interaction of these two variables on reading comprehension skill at retention.

Expected results. It is hypothesized that there will be a main effect of mastery and of Entropy. Thus, students who were transitioned as the result of stealth assessments (either mastery or Entropy) will generate higher quality self-explanations at posttest and will score higher on the retention comprehension measure. These students are the ones who will be receiving information about the optimality of their performance and choices within iSTART-2. Thus, it is hypothesized that their metacognitive awareness may also be improving. It is also hypothesized that there will be a significant interaction between mastery and Entropy on both posttest self-

explanation quality and retention reading comprehension skill. Thus, the influence that these on-line measures have on learning outcomes is most prevalent when both mastery and Entropy thresholds are used to transition students toward more optimal patterns. The students who receive information about both their choice patterns and their skill level may be more likely to monitor both what they do and how they perform within iSTART-2. This behavior would be the most reflective of SRL behavior, thus, it is hypothesized to lead to the greatest learning gains.

Follow-up Study 3 Discussion. The hypotheses in follow-up Study 3 add to our understanding of how online-measure can be used to optimize ITSs. Specifically, these measures can be used to help students guide their learning trajectory. Within the SRL literature, it is stated that many students are unaware of how to set strategic plans during learning tasks (Zimmerman, 2008). Thus, if online measure can automatically detect when students are having difficulty in setting strategic plans and provide subsequent guidance as to what the student should do next, this may prompt metacognitive awareness within the student. Students who receive this guidance may begin to pick up cues from the system and eventually be able to recognize and correct for non-optimal learning patterns. Previous work has shown that with instruction students are able to improve their SRL ability (Muraven, Baumeister, & Tice, 1999). Indeed, the use of online measures may help students become more aware of their behaviors and give them suggestions on how they can optimize their time within the system.

Conclusion

This thesis explored the utility of log-data as a form stealth assessment for students' ability to self-regulate. These analyses approach self-regulation in a novel way by examining nuances in students' log-data to capture tendencies in their choice selections and behaviors across time. The various methodologies presented here 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 seems critical to our understanding of the various ways in which students exert a strategic plan over their learning experience. The three proposed follow-up studies within this thesis lay out how these online-measured can be used and manipulated in the future to provide a better understating of agency, strategic planning and SRL. Overall, these findings and proposed studies may afford researchers the opportunity to further examine the dynamical nature of self-regulation and its impact on learning outcomes.

REFERENCES

- Alexander, P. A., Murphy, P. K., Woods, B. S., Duhon, K. E., & Parker, D. (1997). College instruction and concomitant changes in students' knowledge, interest, and strategy use: A study of domain learning, *Contemporary Educational Psychology*, 22, 125-146.
- Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*, 138, 432-447.
- Arneodo, M., Arvidson, A., Badałek, B., Ballintijn, M., Baum, G., Beaufays, J., & Prytz, K. (1995). Measurement of the proton and the deuteron structure functions. *Physics Letters B*, 364, 107-115.
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the Cognitive Tutor classroom: When students "game the system." *Proceedings of ACM CHI 2004: Computer-Human Interaction* (pp. 383-390). New York: ACM.
- Bandura, A. (2000). Self-efficacy: The foundation of agency. In W. Perig and A. Grob (Eds.). *Control of human behavior, mental processes, and consciousness: Essays in honor of the 60th birthday of August Flammer*, (pp. 17-33) Mahwah, NJ: Erlbaum.
- Bell, C., & McNamara, D. S. (2007). Integrating iSTART into a high school curriculum. In *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 809-814).
- Benhamou, S., & Bovet, P. (1989). How animals use their environment: a new look at kinesis. *Animal Behavior*, 38, 375-383.
- Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, 54(2), 199-231.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, 65(3), 245-281.
- Carroll, J. A. (1963) Model For School Learning. *Teachers College Record*, 64, 723-733.
- Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182.
- Cleary, T. J. (2006). The development and validation of the self-regulation strategy inventory—self-report. *Journal of School Psychology*, 44(4), 307-322.
- Collins, J. J., & De Luca, C. J. (1994). Random walking during quiet standing. *Physical Review Letters*, 73, 764-767.

- 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.
- de Bruin, A. B., Thiede, K. W., & Camp, G. (2001). Generating keywords improves metacomprehension and self-regulation in elementary and middle school children. *Journal of Experimental Child Psychology, 109*, 294-310.
- D’Mello, S. K., Taylor, R., & Graesser, A. C. (2007). Monitoring affective trajectories during complex learning. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 203-208). Austin, TX: Cognitive Science Society.
- Derryberry, D., & Reed, M. A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology, 111*, 225-236.
- Deterding, S., Sicart, M., Nacke, L., O’Hara, K., & Dixon, D. (2011). Gamification: Using game-design elements in non-gaming contexts. In *PART 2 of the Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems* (pp. 2425-2428). ACM.
- 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.
- Facer, K., Joiner, R., Stanton, D., Reid, J., Hull R., & Kirk, D. (2004). Savannah: Mobile gaming and learning? *Journal of Computer Assisted Learning, 20*, 399-409.
- Garner, R., Alexander, P. A., Gillingham, M. G., Kulikowich, J. M., & Brown, R. (1991). Interest and learning from text. *American Educational Research Journal, 28*, 643-659.
- Gillam, B. D., & James, S. R. (1999). U.S. Patent No. 5,964,660. Washington, DC: U.S. Patent and Trademark Office.
- Glaser, C., & Brunstein, J. C. (2007). Improving fourth-grade students' composition skills: Effects of strategy instruction and self-regulation procedures. *Journal of Educational Psychology, 99*, 297-310.
- Godwin, K. E., & Fisher, A. V. (2011). Allocation of Attention in Classroom Environments: Consequences for Learning. *Paper Presented at the Annual Meeting of the Cognitive Science Society*, Boston, Massachusetts.
- Grossman, E. R. F. W. (1953). Entropy and choice time: The effect of frequency unbalance on choice-response. *Quarterly Journal of Experimental Psychology, 5*, 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.
- Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional and cognitive interest. *Journal of Educational Psychology, 89*, 92-102.
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*, 414.
- Harris, K. R., Friedlander, B. D., Saddler, B., Frizzelle, R. & Graham, S. (2005). Self-monitoring of attention versus self-monitoring of academic performance: Effects among students with ADHD in the general education classroom. *Journal of Special Education, 39*, 145-156.
- Hidi, S., & Baird, W. (1986). Interestingness—A neglected variable in discourse processing. *Cognitive Science, 10*, 179-194.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers, 116*, 770-808.
- Jackson, G. T., Boonthum, C., & McNamara, D. S. (2009). iSTART-ME: Situating extended learning within a game-based environment. In H.C. Lane, A. Ogan, and V. Shute (Eds.), *In Proceedings of the Workshop on Intelligent Educational Games at the 14th Annual Conference on Artificial Intelligence in Education*, 59-68. Brighton, UK: AIED.
- Jackson, G.T., Boonthum, C., & McNamara, D.S. (2010). The efficacy of iSTART extended practice: Low ability students catch up. In J. Kay & V. Aleven (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems* (pp. 349-351). Berlin/Heidelberg: Springer.
- Jackson, G. T., Boonthum-Denecke, C. B., & McNamara, D. S. (2012). A comparison of gains between educational games and a traditional ITS. In P. M. McCarthy & G. M. Youngblood (Eds.), *Proceedings of the 25th International Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 444-449). Menlo Park, CA: The AAAI Press.
- Jackson, G.T., Dempsey K.B., & McNamara, D.S. (2010). The evolution of an automated reading strategy tutor: From classroom to a game-enhanced automated system. In M.S. Khine & I.M. Saleh (Eds.), *New Science of learning: Cognition, computers and collaboration in education* (pp. 283-306). New York, NY: Springer.
- Jackson, G.T., Graesser, A.C., & McNamara, D.S. (2009). What students expect may have more impact than what they know or feel. 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. 73-80). Amsterdam, The Netherlands: IOS Press.
- 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.
- Jackson, G. T., & McNamara, D. S. (2012). Applying NLP metrics to students' self explanations. In P.M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 261-275). Hershey, PA: IGI Global
- Jackson, G. T., & McNamara, D. S. (2013) Motivation and Performance in a Game- based Intelligent Tutoring System. *Journal of Educational Psychology*.
- Järvelä, S., & Järvenoja, H. (2011). Socially constructed self-regulated learning and motivation regulation in collaborative learning groups. *Teachers College Record, 113*, 350-374.
- Kim, B., Park, H., & Baek, Y. (2009). Not just fun, but serious strategies: Using meta-cognitive strategies in game-based learning. *Computers & Education, 52*(4), 800-810.
- Landauer, T. K., McNamara, D. S., Dennis, S. E., & Kintsch, W. E. (2007). *Handbook of latent semantic analysis*. Lawrence Erlbaum Associates Publishers.
- Lee, C. S., Goh, D. H. L., Chua, A. Y., & Ang, R. P. (2010). Indagator: Investigating perceived gratifications of an application that blends mobile content sharing with gameplay. *Journal of the American Society for Information Science and Technology, 61*(6), 1244-1257.
- Lobry, J. R. (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biological Evolution, 13*, 660–665.
- Lord, R. G., Diefendorff, J. M., Schmidt, A. M., & Hall, R. J. (2010). Self-regulation at work. *Annual review of psychology, 61*, 543-568.
- Mandelbrot, B. B. (1982). *The fractal geometry of nature*. New York: Freeman.
- McCombs, B. L. (1989). Self-regulated learning and academic achievement: A phenomenological view. In *Self-regulated learning and academic achievement* (pp. 51-82). Springer New York.

- MacGinitie, W. H., & MacGinitie, R. K. (1989). Gates MacGinitie reading tests. Chicago: Riverside.
- Magliano, J. P., Todaro, S., Millis, K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S. (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research*, 32, 185-208.
- McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, 38, 1-30.
- McNamara, D.S. (2009). The importance of teaching reading strategies. *Perspectives on Language and Literacy*, 35, 34-40.
- 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. (2009). Intelligent tutoring and games (ITaG). In *Proceedings of the Workshop on Intelligent Educational Games at the 14th Annual Conference on Artificial Intelligence in Education* (pp. 1-10).
- 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., & 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., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, 34, 147-171.
- 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.
- Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, 10, 98-129.

- 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*, 446-457.
- Nelson, C. R., & Plosser, C. R. (1982). Trends and random walks in macroeconomic time series: some evidence and implications. *Journal of Monetary Economics, 10*, 139-162.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543-578.
- Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, H. E., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E, 49*, 1685-1689.
- Peng, C. K., Havlin, S., Hausdorff, J. M., Mietus, B. S. J., Stanley, H. E., & Goldberger, A. L. (1995). Fractal mechanisms and heart rate dynamics: Long-range correlations and their breakdown with disease. *Journal of Electrophysiology, 28*, 59-65.
- Phillips, L. M., Norris, S. P., Osmond, W. C., & Maynard, A. M. (2002). Relative reading achievement: A longitudinal study of 187 children from first through sixth grades. *Journal of Educational Psychology, 94*, 3-13.
- 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., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33-40.
- Rai, D., & Beck, J. (2012). Math learning environment with game-like elements: An experimental framework. *International Journal of Game Based Learning, 2*, 90- 110.
- Rodrigo, M. M. T., & Baker, R. S. J. D. (2011). Comparing learners' affect while using an intelligent tutor and an educational game. *Research and Practice in Technology Enhanced Learning, 6*, 43-66.
- Rowe, J., McQuiggan, S., Robison, J., & Lester, J. (2009). Off-Task Behavior in Narrative-Centered Learning Environments. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Education, 99-106*.
- Rowe, J., Shore, L., Mott, B., & Lester, J. (2010). Integrating Learning and Engagement in Narrative-Centered Learning Environments. *Proceeding of Intelligent Tutoring Systems, 166-177*.
- Sabourin, J., Rowe, J., Mott, B., & Lester, J. (2011). When off-task is on-task: the affective role of off-task behavior in narrative-centered learning environments. *In Artificial Intelligence in Education* (pp. 534-536). Springer Berlin/Heidelberg.

- 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.
- Schunk, D. H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational psychology review*, *20*, 463-467.
- Schunk, D. H., & Zimmerman, B. J. (2003). Self-regulation and learning. *Handbook of Psychology*.
- 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. *Serious games: Mechanisms and effects*, 295-321.
- Snow, E. L., Jackson, G. T., Varner, L. K., & McNamara, D. S. (2013a). 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, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 272-275). Heidelberg, Berlin, Germany: Springer.
- Snow, E. L., Jackson, G. T., Varner, L. K., & McNamara, D. S. (2013b). The impact of performance orientation on students' interactions and achievements in an ITS. In C. Boonthum-Denecke & G. M. Youngblood (Eds.), *Proceedings of the 26th Annual Florida Artificial Intelligence Research Society (FLAIRS) Conference* (pp. 521-526). Menlo Park, CA: The AAAI Press.
- Snow, E. L., Jackson, G. T., Varner, L. K., & McNamara, D. S. (2013c). The impact of system interactions on motivation and performance. In *Proceedings of the 15th International Conference on Human-Computer Interaction (HCI)* (pp. 103-107). 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.
- Soller, A., & Lesgold, A. (2003). A computational approach to analyzing online knowledge sharing interaction. In *the 11th annual Proceedings of Artificial Intelligence in Education* (pp. 253-260). Sydney, Australia: Springer.
- Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational researcher*, *9*(11), 11-16.

- Tauer, J. M., & Harackiewicz, J. M. (2004). The effects of cooperation and competition on intrinsic motivation and performance. *Journal of Personality and Social Psychology*, 86, 849-861.
- Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Doyne Farmer, J. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1), 77-94.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46, 197-221.
- Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132(3), 331-350.
- Winne, P. H. (2013). Learning Strategies, Study Skills, and Self-Regulated Learning in Postsecondary Education. In *Higher Education: Handbook of Theory and Research* (pp. 377-403). Springer Netherlands.
- Winne, P. H., & Hadwin, A. F. (2013). nStudy: Tracing and supporting self-regulated learning in the Internet. In *International Handbook of Metacognition and Learning Technologies* (pp. 293-308). Springer New York.
- Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review*, 20(4), 429-444.
- Woolf, B. P. (2009). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Burlington MA: Morgan Kaufman Publishers.
- Wolters, C.A. (2011). Regulation of motivation: Contextual and social aspects. *Teachers College Record*, 113, 265-283.
- 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. (1989). *Self-regulated learning and academic achievement: Theory, research, and practice*. Springer-Verlag Publishing.

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.

Zimmerman, B. J., & Schunk, D. H. (2013). *Self-regulated learning and academic achievement: Theoretical perspectives*. Routledge.