

Taking Control: Stealth Assessment of Deterministic Behaviors Within a Game-Based System

Erica L. Snow¹ · Aaron D. Likens¹ ·
Laura K. Allen¹ · Danielle S. McNamara¹

Published online: 7 December 2015

© International Artificial Intelligence in Education Society 2015

Abstract Game-based environments frequently afford students the opportunity to exert agency over their learning paths by making various choices within the environment. The combination of log data from these systems and dynamic methodologies may serve as a stealth means to assess how students behave (i.e., deterministic or random) within these learning environments. The current work captures variations in students' behavior patterns by employing two dynamic 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 manner also generated higher quality self-explanations during training sessions. The results point toward the potential for dynamic analyses such as random walk analyses and Hurst exponents to provide stealth assessments of students' learning behaviors while engaged within a game-based environment.

Keywords Game-based intelligent tutoring systems · Dynamic analyses · Stealth assessment · Log data

Game-based Intelligent Tutoring Systems (ITSS) are computer-based learning environments that provide students with pedagogical instruction within the context of a game

✉ Erica L. Snow
Erica.L.Snow@asu.edu

Aaron D. Likens
Aaron.Likens@asu.edu

Laura K. Allen
LauraKAllen@asu.edu

Danielle S. McNamara
Danielle.McNamara@asu.edu

¹ Department of Psychology, Arizona State University, Tempe, AZ 85287, USA

(Van Eck 2007). Game-based ITSs can be situated in a variety of domains such as science (Johnson-Glenberg et al. 2011; Sabourin et al. 2012), mathematics (Rai and Beck 2012), and technology education (van Eck 2006). One key feature of these environments is that they often afford students the opportunity to exert agency over their learning path by allowing for multiple methods and trajectories of play (King and Cazessus 2014; Sabourin et al. 2012; Schmierbach et al. 2012; Teng 2010). This inevitably leads to students interacting with and experiencing the game-based environment differently. For example, examining the ways in which students behave when they are afforded this agency can lead to a better understanding of optimal and non-optimal behaviors within a learning environment (Sabourin et al. 2012). The inclusion of agentic features (e.g., choose your own path or edit an avatar) has been associated with increased immersion, motivation, and positive learning gains (Cordova and Lepper 1996; Schmierbach et al. 2012; Teng 2010).

Although variations in students' behaviors may prove to be invaluable information for researchers, these behaviors are often difficult to measure and quantify. Traditionally, scientists have used self-report measures as proxies to gauge students' actions and behaviors during learning tasks (Rosenbaum 1980; Zimmerman and Schunk 1989, 2001; Zimmerman 1990). While informative, traditional self-report measures that assess students' behaviors and intentions during learning tasks may not fully or adequately capture their target construct (Hadwin et al. 2007; Zhou 2013). Indeed, an overarching concern regarding self-reports is 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 of what comprises good and poor performance, leading to over or under estimations of various traits. Third, the behaviors, cognitive states, and affect 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, learning strategies and behaviors are dynamic (Hadwin et al. 2007; Lord et al. 2010). 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 develop and change over time. Hence, static measures may not adequately capture nuanced changes in how learners modulate and change their behaviors across varying goals and task demands.

Online Measures

Online measures offer an alternative means of capturing the dynamic nature of learning behaviors (Hadwin et al. 2007; Ventura and Shute 2013; Winne and Hadwin 2013; Zhou 2013). In contrast to offline self-report and post assessments, online measures capture behaviors from the learner in real-time. These measures capture nuanced patterns in students' behaviors and thus may be more likely to capture how students exert agency while engaging in learning tasks.

Within automated learning environments, online measures such as log data can act as a form of stealth assessment by unobtrusively capturing variations in students' behaviors (Shute et al. 2009; Shute 2011). Log data, also referred to as keystroke, mouse click, click stream, or telemetry data (depending on the context), 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 (Hadwin et al. 2007; Sabourin et al. 2012; Schulte-Mecklenbeck et al. 2011; Shih et al. 2010; Snow et al. 2014).

For instance, Hadwin and colleagues (Hadwin et al. 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 and colleagues examined how students' patterns of annotation and study habits informed 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 (Sabourin et al. 2012) examined how log data from the narrative-centered environment, Crystal Island, was indicative of students' 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 strategy use. They found that the inclusion of system log data significantly contributed to the classification of students' use of metacognitive strategies. Such research demonstrates that log data extracted from adaptive environments yield unique and unobtrusive means to examine the ways in which individuals behave during learning tasks, and these behaviors are important indicators of individual differences that contribute to learning outcomes.

Dynamic Systems Theory

In conjunction with log data, dynamic systems theory and its associated analysis techniques offer researchers a unique 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. Dynamic 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.

Dynamic methodologies have been utilized in adaptive systems to investigate the complex patterns that emerge in students' behaviors (Hadwin et al. 2007; Snow et al. 2013; Soller and 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 choice patterns that manifest within students' log data are likely to be overlooked using more traditional (e.g., static) statistical analyses, and that dynamic analyses offer a readily available means to capture this crucial source of information.

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 dynamic 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 and 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 and Plosser 1982), ecology (Benhamou and Bovet 1989), psychology (Collins and De Luca 1994), 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 adaptive systems vary as a function of individual differences (Snow et al. 2013).

While random walk analyses generate visualizations of unique 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 (i.e., deterministic) quality. Time series fluctuations exhibiting deterministic tendencies are assumed to reflect self-organized and controlled processes (Van Orden et al. 2003). By contrast, when each moment in a time series is independent of every other moment, the fluctuations in the time 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 and De Luca 1994). Time series fluctuations exhibiting antipersistent behaviors are assumed to be demonstrating corrective processes (Collins and De Luca 1994).

The goal of the current study is to investigate how variations in students' behaviors manifest across time, and ultimately impact daily learning outcomes within a game-based system. Random walks and Hurst exponents are used to capture the fine-grained behavior patterns that manifest within students' log data collected across multiple sessions within a complex learning environment. Ultimately, the combination of log data and dynamic techniques may serve as novel forms of stealth assessment, examining students' propensity to act in deterministic (or random) manners across time, and 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 and McNamara 2013). This system provides students with instruction on the use of self-explanation and comprehension strategies (Jackson et al. 2012; Jackson and McNamara 2013). iSTART-ME is an ideal environment to examine how patterns manifest in students' choices across time 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 et al. 2004) but integrates games and game-based features to enhance students' motivation, engagement, and persistence over time (Jackson et al. 2009; Jackson and 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 et al. 1997; Bandura 2000; Pajares 1996; Pintrich 2000; Zimmerman and Schunk 2001). Previous work has revealed that when game-based features are embedded within iSTART-ME, students report an increase in engagement and motivation across multiple training sessions (Jackson and McNamara 2013). The current study takes this work a step farther by examining *how* students interact with these game-based features incorporated within the system interface.

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 et al. 2009). 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 Fig. 1). Additionally, within this selection menu, students can view their current level and the number of skill points and trophies earned.

Spend my iBucks to:		Play a Mini-Game	
Change the Interface	Cost		Cost
Change background color	300 iB	Strategy Match	300 iB
Swap out the tutor	300 iB	Bridge Builder	300 iB
		Vocabulous	300 iB
Edit my character	300 iB	SE Lifeline	300 iB
(shirt, hair, headgear, glasses)		Balloon Bust	300 iB
		Dungeon Escape	300 iB

Fig. 1 Screenshot of the 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 (see Fig. 2). In Map Conquest, students generate self-explanations to earn dice, which are used to conquer squares on a map (see Fig. 3). 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.

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

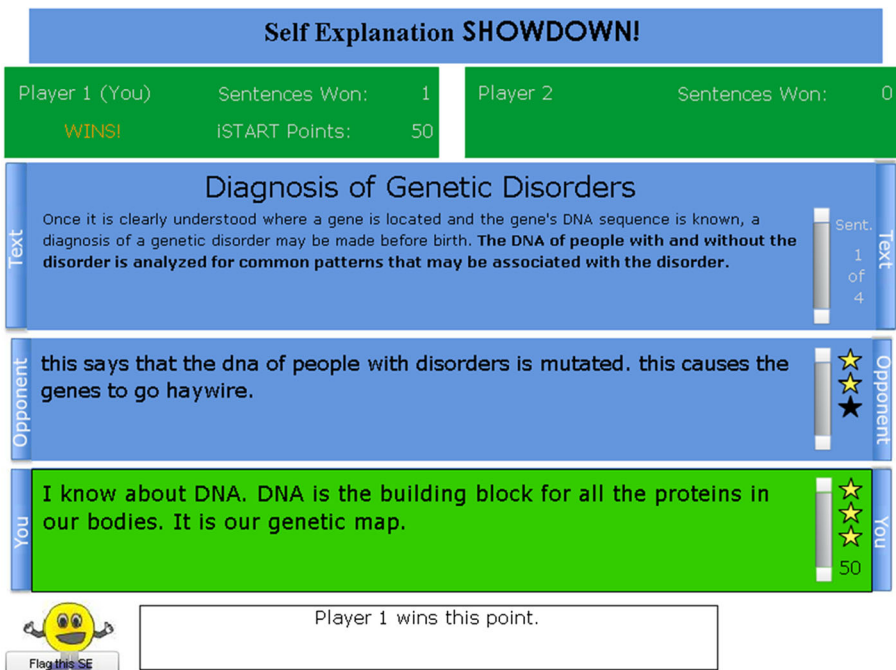


Fig. 2 Screenshot of Showdown



Fig. 3 Screenshot of Map Conquest

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 Fig. 4).

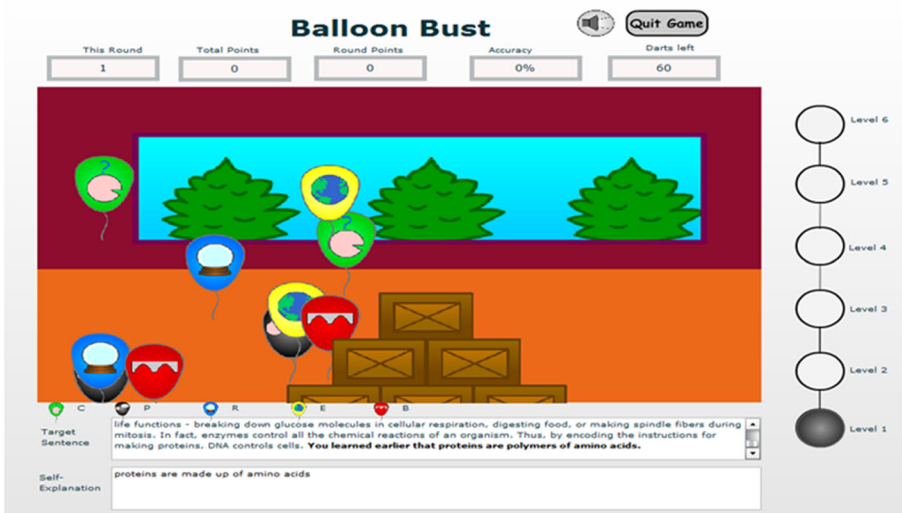


Fig. 4 Screenshot of Balloon Bust

Current Study

In summary, game-based environments afford students multiple methods of interaction and play. Log data from these environments can capture variations in these behaviors to help scientists decipher various learning patterns. In particular, researchers can apply dynamic analyses that focus on the fluid changes in nuanced behavior patterns to gain a deeper understanding of how students control their behaviors over time and ultimately, the impact those behaviors have on learning outcomes. The current study uses two statistical techniques: random walks and Hurst exponents. The combination of these two techniques provides a means to visualize and categorize fine-grained patterns in students' behaviors that emerge within system log data across time. The current study uses these methodologies to examine two research questions.

- 1) Do students demonstrate controlled patterns of interaction (i.e., deterministic) within the game-based system iSTART-ME?
- 2) How do variations in students' interaction patterns impact in-system performance, posttest, and long-term retention learning outcomes (i.e., self-explanation quality)?

Method

Subjects

The data that are analyzed in this paper were collected as part of a larger laboratory study that compared three conditions: iSTART-ME, iSTART-Regular, and a no-tutoring control (Jackson and 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. Of the 40 students, 50 % were female, and 17 % were Caucasian, 73 % were African-American, and 10 % were other nationalities. All participants were monetarily compensated for their participation.

Procedure

The study comprised 11 sessions within a laboratory experiment 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 for approximately 1 h per session. 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 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 et al. 2007) and word-based measures (McNamara et al. 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 that relate to the target sentence, but lack elaborations that use 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). Using this algorithm, students' self-explanations were scored at pretest, training, posttest, and retention. Within the current study, students' training self-explanation scores were averaged (across all 8 sessions) to create an aggregate score that represented their overall performance within the system.

System Interactions

Students interacted freely within iSTART-ME for 8 training sessions. Every choice a student made was logged into the system database. We then categorized every interaction choice within those raw data files into one of four game-based categories (described below). It is important to note that only completed actions were retained for this analysis. Thus, if a student opened a game and then exited the game without finishing it (regardless of time spent in the game), that interaction would not be counted as a game-played and therefore would not be included in the final analyses. The analysis included a total of 11,120 game-based interactions, with an average of 278 ($SD = 33$) choices per student.

As described earlier, students' interactions with iSTART-ME involved one of four types of game-based features, each representing a different type of game-based functionality within the system:

1. *Generative practice games.* iSTART-ME includes three practice environments (Coached Practice, Map Conquest, and Showdown) that prompt students to generate their own self-explanations within the context of a game. Within the generative practice environments students receive feedback concerning their self-explanations. Thus, generative practice games are designed to provide students with opportunities to apply comprehension strategies while reading challenging texts, and receive feedback on the quality of their self-explanations. On average, students interacted with generative practice games 19.03 times ($SD = 7.07$).

2. *Identification mini-games.* There are six identification mini-games that reinforce the learning strategies and goals presented by asking the students to identify the type of strategies used within example self-explanations. These games do not prompt students to generate their own self-explanation, but instead provide students with strategy recognition practice. This involves students reading the text and an explanation of the text, and then choosing the principal strategy used to generate that explanation. On average, students interacted with the identification mini-games 24.9 times ($SD = 29.17$).
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. The personalizable features potentially provide a means to enhance students' engagement and afford a feeling of personal investment within the game interface (Jackson and McNamara 2013). However, they also potentially distract from the learning process because they are unrelated to learning how to better understand challenging text. On average, students interacted with personalizable features 6 times ($SD = 8.04$).
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. Achievement screens were embedded within the system to assess the relation between monitoring these sources of information about performance and the learning outcomes. For example, if students are able to track their progress within a system, they in turn may become more personally invested in their performance. Alternatively, tracking this information may distract students from the learning process. On average, students interacted with achievement screens 45.45 times ($SD = 36.09$).

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. 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 $(0,-1)$. These locations are random and are not associated with any qualitative value associated with the activity.

Each student's unique 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. Thus, the directionality of the axes can vary as long as they are consistent throughout the entire analyses. In the current work, the axis direction assignment was set prior to the analysis and remained consistent for every student.

Figure 5 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 Fig. 5). 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 Fig. 5). 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 Fig. 5). The student's third interaction was with an achievement screen, which moves the particle one unit down along the Y-axis (see # 3 in Fig. 5). The fourth hypothetical interaction was with another achievement screen, which again moves the particle one unit down along the Y-axis (see # 4 in Fig. 5). 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 Fig. 5). These simple rules were utilized for every interaction a student made within iSTART-ME. This analysis resulted in a unique walk for each of the 40 students.

Figure 6a and b illustrate two random walks that were generated using students' log data. In Fig. 6a (random walk on the left), the generated random walk reveals that this particular student interacted most frequently with the generative practice games. This walk trajectory is primarily anchored along the generative practice axis. Conversely, the student who generated the random walk in Fig. 6b (random walk on the right) 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 games and identification mini-games axes. These two contrasting figures demonstrate how log data can be used to generate a unique spatial representation of each student's time in the system. Random walks provide a visualization of students' interaction paths within the iSTART-ME system. It is important to note that this technique can be used on any number of categorical variables. In the current study, all walks lay on an XY axis; however, these flexible visualization tools can be used on an unlimited number of dimensions and vectors.

Table 1 System interaction choices 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)

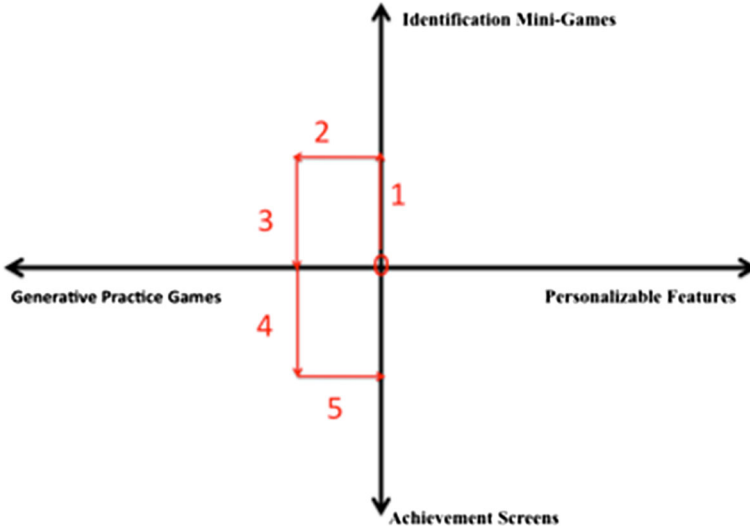


Fig. 5 Example random walk with five interaction choices

Although random walks provide an illustration of students’ movements within a game-based environment, they do not provide a quantification of these patterns. Thus, to quantify patterns of movements from these walks, distance time 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:

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

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

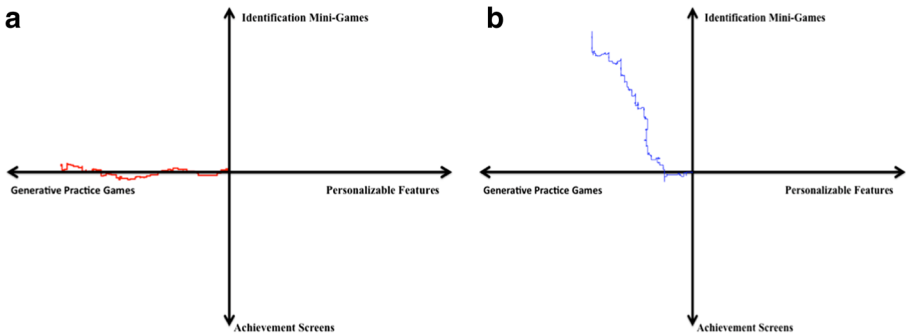


Fig. 6 a and b. Random walks generated from two different students’ log data

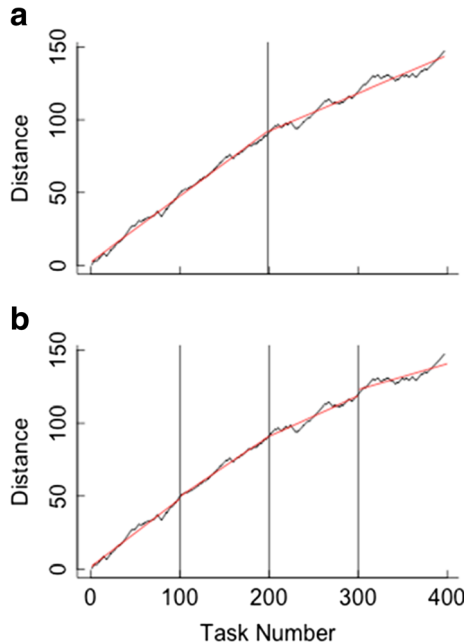


Fig. 7 Illustration of the first (a) and second (b) iteration of the second step of the DFA procedure for a single time series. In both figures, vertical lines represent the binning procedure that occurs during the second step of DFA

(DFA; Peng et al. 1994). DFA is a method for estimating persistence (i.e., deterministic tendencies) in a time series by determining how a measure of variance depends on scale size (Peng et al. 1994). The DFA algorithm is captured by the following equation:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2} \tag{2}$$

where N is the total number of observations, $y(k)$ is the k th observation, $y_n(k)$ is the predicted value of $y(k)$ from a local trend line, and n is the window size for a given scale. More concretely, the DFA algorithm involves four simple steps. The first step is to create the *profile* by subtracting the mean from the time series and then taking the cumulative sum (i.e., integrating). The second step involves dividing a series of length, N , into N/n non-overlapping bins, such that each bin contains n observations. The third step is to compute the root-mean-square residual across all bins (Fig. 7a). The residual is obtained by subtracting local trend lines within each bin, this process is repeated for several n s, decreasing n by a power of 2 (Fig. 7b). The maximum n should be $N/2$, and the result is a fluctuation function, $F(n)$. The fourth and final step is to regress $\log_2 F(n)$ on $\log_2 n$ (Fig. 8). In the case of persistence, the expected result from the final step is a linear slope, α , greater than 0.5.

The above steps are depicted in Figs. 7 and 8. Figure 7 shows an example time series for a single student’s interaction trajectory, where each interaction is taken as the analogue of a unit of time. The precedence for using observations in that manner can be found throughout the literature (e.g., Peng et al. 1994; Van Orden et al. 2003).

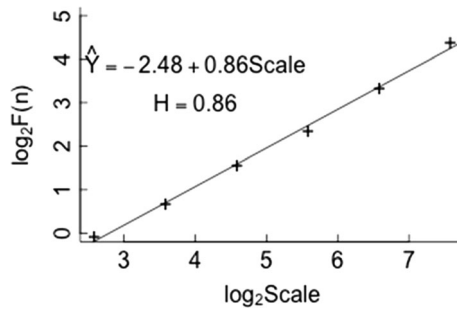


Fig. 8 Depiction of the fourth and final step in detrended fluctuation analysis. The equation represents the regression procedure, and the slope indicates the Hurst exponent

Figure 7 also depicts two of the detrending steps that give DFA its name. Figure 7a shows the first of those steps: a regression line is fit to each of the demarcated bins. The fitted lines are subtracted from the time series to obtain the residuals used to compute $F(n)$ at a given scale. Figure 7b displays the subsequent iteration by reducing the window sizes by exactly one-half and repeating the fitting and detrending procedure. Once $F(n)$ has been obtained for all n , the regression analysis depicted in Fig. 8 is conducted. That is, the base two logarithm of $F(n)$ is regressed onto the base two logarithm of Scale (i.e., n). The resulting slope is the Hurst exponent. The interpretive index for Hurst is as follows: $0.5 < H \leq 1$ indicates persistent (deterministic or controlled) behavior, $H = 0.5$ signifies random behavior, and $0 \leq H < 0.5$ denotes antipersistent behavior.

Within the current work, Hurst exponents are used to quantify changes in students' interaction patterns (i.e., distance time series). We are specifically interested in whether students' choice patterns reveal deterministic (i.e., controlled) or random (i.e., independent) tendencies. Hurst exponents act as long-term correlations; therefore they provide a metric about how an entire interaction pattern changes and manifests across time. In the current work, deterministic interaction patterns are assumed to reflect self-organized and controlled processes (Van Orden et al. 2003). By contrast, interaction patterns that exhibit random tendencies reflect a breakdown in system functioning and control (e.g., Peng et al. 1995). Finally, when interaction patterns exhibit antipersistent tendencies, they are acting as corrective (i.e., negatively correlated) processes (Collins and De Luca 1994). Using this classification, the Hurst exponent affords us the opportunity to examine how controlled students are when they interact with the four types of game-based features embedded within iSTART-ME (i.e., generative practice games, identification mini-games, personalizable features, and achievement screens).

Surrogate Analysis

Surrogate analysis is an important step in time series analysis when using brief time series, as in the present case (Theiler et al. 1992). Clearly, the DFA procedure outlined above could be applied to any time series and result in a Hurst exponent. What is needed, though, is a means to determine whether the observed exponent accurately represents the underlying process. Surrogate analysis fills that need by providing a principled, statistical means to distinguish time series generated by random processes

from time series generated by deterministic processes. The general approach of surrogate analysis is to compare an observed measure—like the Hurst exponent—to similar measures derived from randomly shuffled surrogate data (Theiler et al. 1992). The idea is that the analyzed time series may be a random process that merely appears to exhibit persistent- or antipersistent-like behavior over a short interval. If so, then randomly shuffling the series should not affect the scaling structure. If not, and the scaling behavior is genuine, then shuffling the time series should deteriorate the scaling structure. Surrogate analysis tests those hypotheses.

In the current context, the surrogate analysis tests the null hypothesis that the observed Hurst exponent is an artifact of short series length. We implemented the surrogate analysis by shuffling each time series 40 times and then by performing a DFA on each of the shuffled series. We then compared the average surrogate derived Hurst exponent for each time series with its observed Hurst exponent counterpart using a paired samples *t*-test. The test revealed that shuffled surrogates produced smaller Hurst exponents than did the intact series, $t(39) = 156.90$, $p < 0.001$, supporting the conclusion that the observed Hurst exponents accurately represent the observed patterns of persistence.

Results

Hurst Exponents and Surrogate Analyses

To characterize how students interacted with the system over the course of the 8 training sessions, 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 (some random tendencies) 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, $t(39) = 156.90$, $p < 0.001$, suggesting that the Hurst exponents characterizing students' interaction patterns were reliable.

High and Low Hurst Student Examples

Within the current study, students' Hurst exponents ranged considerably. Indeed, some students acted in a weakly persistent manner, while others were more deterministic in their choice patterns. Figure 9 illustrates how *two* students (one low Hurst and one high Hurst) differed from each other in terms of percentage of interactions with each type of game-based feature. In Fig. 9, the low Hurst student (Hurst of 0.60) demonstrated more variation in the interaction pattern, where no one feature was favored. Conversely, a high Hurst student (Hurst of 0.90) interacted primarily with the generative practice games. While these are only two examples of the difference between high and low Hurst students, the general notion is that low Hurst students acted more impetuously and jumped around more frequently, whereas the high Hurst students acted in more controlled and deterministic manners.

Interaction Choices

Figure 9 demonstrates how high and low Hurst students' choices of interactions varied between two students. However, to examine relations 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 (i.e., Hurst exponents) within the system were not related to any specific feature. Thus, Hurst exponents are not capturing *what* game-based features students interact with, but rather *how* they interact with these features.

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 (in-system performance), 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 their average 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 are consistent with previous work showing that for generative activities, such as self-explanation, the impact of training is not observed immediately after training (at posttest). Rather, the effects of training are more likely to be apparent after a delay (e.g., Adams et al. 2014; Dunlosky et al. 2013; Schmidt and Bjork 1992). Overall, the results from this analysis indicate that when students'

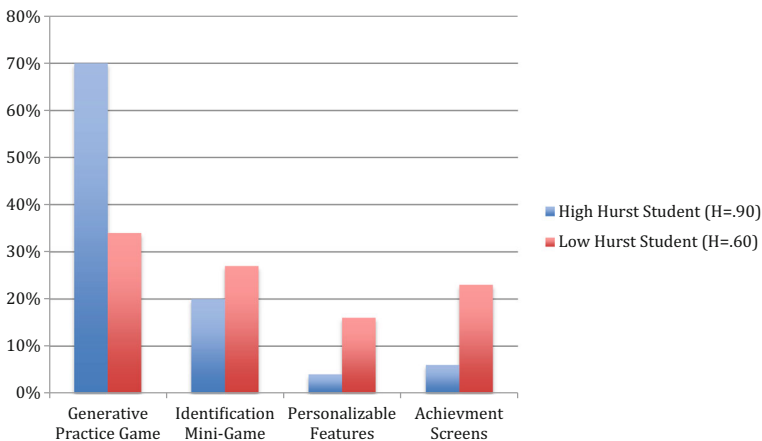


Fig. 9 Two sample students' percentage of game-based interactions as a function of Hurst score, where high Hurst is indicative of more deterministic behavior patterns

interaction patterns within the system reveal more deterministic properties, they generated higher quality self-explanations during training and at retention.

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

A similar 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).

Discussion

Game-based environments frequently afford students opportunities to exert agency over their learning path. A predominant assumption by many researchers and educators is that students' ability to control 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, assessing variations in these behaviors can be difficult, and traditionally has relied 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 dynamic measures are needed to gain a deeper understanding of students' ability to control their

Table 2 Correlations between self-explanation scores and Hurst exponents

Strategy Performance	<i>r</i>
Training self-explanation score	0.51**
Posttest self-explanation score	0.09
Retention self-explanation score	0.31*

$p = .05^*$, $p < .001^{**}$

Table 3 Linear regression analyses predicting daily self-explanation quality

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

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

behaviors (Hadwin et al. 2007). Log data has been previously used to analyze variations in students' learning behaviors at more fine-grained levels (Hadwin et al. 2007; Sabourin et al. 2012; Snow et al. 2014). The work presented here builds upon these findings by conducting dynamic analyses of system log data to investigate the extent to which students' behaviors exhibit deterministic or random properties. These initial analyses explore how dynamic 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.

The current study made use of novel methodologies by employing random walk and Hurst exponent analyses in an attempt to capture each student's unique 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' interaction patterns are controlled and deterministic (Mandelbrot 1982; Van Orden et al. 2003). Specifically, when students' interaction patterns have deterministic tendencies, it may be indicative that they are exhibiting persistent and controlled behavior patterns. Conversely, when students' interaction patterns are weakly persistent (indicating random tendencies), it may be indicative that they are *not* behaving with purpose, control, or persistence. These tendencies across long periods of time may reveal trends in how students approach learning tasks. Therefore, this work begins to shed light upon the dynamic nature of learning behaviors that students exhibit while interacting within a game-based environments.

Results from the current study fall in line with previous work that has shown that students' ability to control and regulate their learning behaviors has a positive impact on learning outcomes (Butler and Winne 1995; Pintrich and De Groot 1990; Zimmerman and

Table 4 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	

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

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). These results are consistent with previous work showing that when students engage in generative activities (e.g., self-explanation), the effects on learning are often not apparent immediately after training at posttest, but rather emerge more strongly at delayed retention tests (e.g., Adams et al. 2014; Dunlosky et al. 2013; Schmidt and Bjork 1992).

Overall, these results suggest that when students are given more control over their environment there are some potential consequences, at least for in-system performance. This may be especially important within game-based environments, as they often offer students numerous opportunities to control their trajectory within the system (King and Cazessus 2014; Sabourin et al. 2012; Teng 2010). It is also important to note that within iSTART-ME, it seems to be critical for students to engage within the system in a more deterministic way. However, such agency may or may not be appropriate depending on the learning goals embedded within the environment. In other game-based systems, it may be more prudent for students to explore the system interface more frequently, thus revealing a more impetuous behavior pattern. As such, the findings presented here are meant to provide evidence that students' behaviors within game-based environments are linked to learning outcomes and should be measured by researchers when they evaluate the effectiveness of their respective system.

This study serves as a starting point for scientists to apply dynamic 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 dynamic methodologies show strong promise in providing online stealth indicators of controlled behaviors. The analyses in this study focused on students' movements between four game-based features. However, dynamic methodologies are flexible and the only notable limitation of these methods is that they must be applied to temporal data. Thus, dynamic techniques are generalizable to almost any type of time-stamped log data from any type of system with any number of actions (i.e., choices or behaviors).

One limitation of the current study is that to reliably calculate Hurst exponents, numerous data points are needed. In the current study, over 11,000 game-based interactions were captured, with each student averaging over 275 choices. Thus, as can be imagined, replication of such an in-depth data set is difficult. One way to counter this problem is to use another measure of order or disorder that requires fewer temporal data points. For instance, an Entropy calculation can be used to calculate order and disorder with far fewer data points than the Hurst (Snow et al. 2015). Thus, although Hurst exponents require numerous data points, alternative dynamic methodologies may also capture the degree to which patterns are ordered versus random while using less data.

Another limitation of the current work is that we did not include self-report measures of self-regulation or related constructs. While we hypothesize that students who act in a more deterministic manner are exerting control and may be self-regulating, the results presented here do not support such an assertion. Accordingly, the next steps in this research agenda would 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 controlled behaviors and constructs such as self-regulation. The outcomes of such assessments can then be compared to those provided by dynamic 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. Recently, researchers have begun to move away from the use of self-reports and instead have relied on game-play or performance data to measure constructs such as persistence and deterministic behavior (DiCerbo 2014; Ventura and Shute 2013; Ventura et al. 2013). Thus, the horizon of future research seems to be pointing toward the establishment of the respective utility of using static versus dynamic assessments of learning behaviors.

Another future direction of the current study regards the practical use of this approach. Ultimately, the purpose of using dynamic measures is to capture variations in learning behaviors 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 dynamic techniques as a means to unobtrusively assess students' 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 deterministic trajectory.

In conclusion, this study explored the use of two dynamic methodologies to unobtrusively assess deterministic behaviors and their impact on learning within a game-based environment. These analyses are among the first attempts to examine variations 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 dynamic nature of learning behaviors and their impact on student learning.

Acknowledgments This research was supported in part by the Institute for Educational Sciences (IES R305G020018-02, R305G040046, R305A080589) and National Science Foundation (NSF REC0241144, IIS-0735682). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES or NSF. We are particularly grateful to G. Tanner Jackson who collected the data for this study. We are also grateful to the many other members of the SoLET Lab who have contributed to the completion of this study and this project.

References

- Adams, D. M., McLaren, B. M., Durkin, K., Mayer, R. E., Rittle-Johnson, B., Isotani, S., et al. (2014). Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior*, *36*, 401–411.
- 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.
- Ameodo, M., Arvidson, A., Badelek, B., Ballintijn, M., Baum, G., Beaufays, J., et al. (1995). Measurement of the proton and the deuteron structure functions. *Physics Letters B*, *364*, 107–115.
- Bandura, A. (2000). Self-efficacy: the foundation of agency. In W. Perig, & 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.

- Benhamou, S., & Bovet, P. (1989). How animals use their environment: a new look at kinesis. *Animal Behavior*, *38*, 375–383.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, *65*(3), 245–281.
- 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.
- DiCerbo, K. E. (2014). Game-based assessment of persistence. *Journal of Educational Technology & Society*, *17*(1), 17–28.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58.
- 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.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, *116*, 770–808.
- Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology*, *105*, 1036–1049.
- Jackson, G. T., Boonthum, C., & McNamara, D. S. (2009). ISTART-ME: situating extended learning within a game-based environment. In H. C. Lane, A. Ogan, & V. Shute (Eds.), *In proceedings of the workshop on intelligent educational games at the 14th annual conference on artificial intelligence in education* (pp. 59–68). Brighton, : AIED.
- 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, : The AAAI Press
- Johnson-Glenberg, M. C., Birchfield, D., Savvides, P., & Megowan-Romanowicz, C. (2011). *Semi-virtual embodied learning-real world stem assessment. In serious educational game assessment* (pp. 241–257). Rotterdam: Sense Publishers.
- King, C., & Cazessus, M. (2014). Teaching With AudaCity: A Board Game for Urban Studies. In *ECGBL2014-8th European Conference on Games Based Learning: ECGBL2014* (p. 272). Academic Conferences and Publishing International
- Landauer, T. K., McNamara, D. S., Dennis, S. E., & Kintsch, W. E. (2007). *Handbook of latent semantic analysis*. Mahwah: Lawrence Erlbaum Associates Publishers.
- 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.
- McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: challenges and successes. *Megacognition and Learning*, *3*, 1–11.
- 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., 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.
- 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.
- 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.
- Rosenbaum, M. (1980). A schedule for assessing self-control behaviors: preliminary findings. *Behavior Therapy*, 11(1), 109–121.
- 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). Berlin: Springer.
- Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: common principles in three paradigms suggest new concepts for training. *Psychological Science*, 3, 207–217.
- Schmierbach, M., Limperos, A. M., & Woolley, J. K. (2012). Feeling the need for (personalized) speed: how natural controls and customization contribute to enjoyment of a racing game through enhanced immersion. *Cyberpsychology, Behavior, and Social Networking*, 15(7), 364–369.
- Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (2011). The role of process data in the development and testing of process models of judgment and decision making. *Judgment and Decision Making*, 6(8), 733–739.
- Shih, B., Koedinger, K. R., & Scheines, R. (2010). Discovery of Student Strategies using Hidden Markov Model Clustering. In the Proceedings of the 6th International Conference on Educational Data Mining
- 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*, 1, 295–321.
- 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,: Springer.
- Snow, E. L., Jackson, G. T., & McNamara, D. S. (2014). Emergent behaviors in computer-based learning environments: computational signals of catching up. *Computers in Human Behavior*, 41, 62–70.
- Snow, E. L., Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Does agency matter?: exploring the impact of controlled behaviors within a game-based environment. *Computers & Education*, 26, 378–392.
- 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,: Springer
- Teng, C. I. (2010). Customization, immersion satisfaction, and online gamer loyalty. *Computers in Human Behavior*, 26, 1547–1554.
- 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.
- Van Eck, R. (2006). Using games to promote girls' positive attitudes toward technology. *Innovate: Journal of Online Education*, 2(3), 6–25.
- Van Eck, R. (2007). Building Artificially Intelligent Learning Games. In D. Gibson, C. Aldrich & M. Prensky (Eds.), *Games and Simulations in Online Learning: Research and Development Frameworks*. Hershey, PA: Information Science, 271–307.
- 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.
- Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior*, 29(6), 2568–2572.
- Ventura, M., Shute, V., & Zhao, W. (2013). The relationship between video game use and a performance-based measure of persistence. *Computers & Education*, 60(1), 52–58.
- 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). New York: Springer.
- 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., & 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.