Computers & Education 86 (2015) 224–235

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game

Valerie J. Shute [a,*], Sidney D'Mello [b], Ryan Baker [c], Kyunghwa Cho [a], Nigel Bosch [b], Jaclyn Ocumpaugh [c], Matthew Ventura [a], Victoria Almeda [c]

[a] Florida State University, USA
[b] University of Notre Dame, USA
[c] Teachers College, Columbia University, USA

ARTICLE INFO

Article history:
Received 19 June 2015
Received in revised form 28 July 2015
Accepted 3 August 2015
Available online 8 August 2015

Keywords:
Affective states
Learning
Physics
Persistence
Engagement

ABSTRACT

This study investigated the relationships among incoming knowledge, persistence, affective states, in-game progress, and consequently learning outcomes for students using the game Physics Playground. We used structural equation modeling to examine these relations. We tested three models, obtaining a model with good fit to the data. We found evidence that both the pretest and the in-game measure of student performance significantly predicted learning outcome, while the in-game measure of performance was predicted by pretest data, frustration, and engaged concentration. Moreover, we found evidence for two indirect paths from engaged concentration and frustration to learning, via the in-game progress measure. We discuss the importance of these findings, and consider viable next steps concerning the design of effective learning supports within game environments.

1. Introduction

Good teachers are keenly aware that students differ along a number of important dimensions which can influence learning. The process of learning new knowledge and skills can trigger a range of emotional responses, wide variation in students' behaviors, and consequently varying learning outcomes. While one teacher cannot manage all of these individual differences in a typical classroom setting, some educational games are beginning to model—with the goal to support—such variation (e.g., Conati & Maclaren, 2009). The main goal of these adaptive educational games is to create an engaging and flexible environment that supports learning for a broad range of learners. Accomplishing this goal depends largely on accurately measuring relevant learner characteristics, such as the type and level of knowledge, skills, personality traits, as well as dynamic cognitive and affective states —and then determining how to leverage the information to improve student learning (Conati, 2002; Park & Lee, 2004; Shute & Zapata-Rivera, 2012; Shute, Lajoie, & Gluck, 2000; Snow, 1994). An additional challenge involves doing all this within the context of a game without disrupting flow (Csikszentmihalyi, 1990), which is often experienced while interacting with a well-designed game.

* Corresponding author.
E-mail address: vshute@fsu.edu (V.J. Shute).

http://dx.doi.org/10.1016/j.compedu.2015.08.001
0360-1315/

This paper describes the results from a multi-method study, which assessed learners along multiple dimensions, including field observations of student affect, measures of persistence and gameplay, and learning (i.e., conceptual physics understanding), within an educational game called Physics Playground (see Shute & Ventura, 2013). The primary aim of our research is to establish the ways that specific affective states (e.g., frustration, confusion boredom, and engaged concentration), persistence, and in-game performance collectively influence learning. We use a structural equation modeling framework to investigate how these factors interact to influence each other and ultimately learning. While individual pairs of these measures have previously been researched, this paper represents – to the best of our knowledge – the first attempt to integrate all of these factors into a comprehensive model.

We chose to use an educational game for our assessment and learning environment for several reasons. Educational games have emerged as a genre of technology that has particularly high potential for creating rich and engaging learning experiences that capture students' enthusiasm and promote meaningful learning (see Clark, Tanner-Smith, & Killingsworth, 2014 for a recent meta-analysis). That is, core features of well-designed games (e.g., problem solving, adaptive challenges, and ongoing feedback) can engender motivation, which in turn supports engagement and learning (e.g., Shute, Rieber, & Van Eck, 2011). In addition, adaptive challenges and dynamic performance feedback in a game help to create an environment that can foster the sense of flow (Csikszentmihalyi, 1990) and potentially cultivate the growth mindset that engenders effort-driven, challenge-centered competency development (Dweck, 2006; Yeager & Dweck, 2012). These same good-game features can also potentially influence persistence (or "learned industriousness," see Eisenberger, 1992) where individuals who are required to exert high effort in one task will continue to exert high effort in a subsequent task (Shute, Ventura, & Ke, 2015). Finally, games allow us to use performance-based assessments of constructs like persistence and engagement which can be more authentic and valid than their self-report measure counterparts (Duckworth & Yeager, 2015; Ventura & Shute, 2013).

Well-designed games can also promote meaningful affective experiences for students, which is important given the inevitable role that affect plays during learning (e.g., Calvo & D'Mello, 2011; Kim & Pekrun, 2014). This is particularly important because affect can indirectly influence learning outcomes by modulating cognitive processes in significant ways (see Fiedler & Beier, 2014). Positive affective states such as delight, excitement, and eureka are experienced when tasks are completed, challenges are conquered, insights are unveiled, and major discoveries are made via creative exploration and problem solving. However, not all affective states experienced in good games are necessarily positively experienced. For example, students get confused when outcomes do not match expectations, when they encounter challenging impasses, and when they are unsure about how to proceed (e.g., Andres et al., 2014; D'Mello & Graesser, 2014a). Frustration occurs when students make mistakes, get stuck, or when important goals are blocked (Kapoor, Burleson, & Picard, 2007). Gee has noted that frustration is a characteristic aspect of many games, even highly successful games; though in a game context, even frustration can be part of a challenging and enjoyable overall experience (Gee, 2007).

Mild levels of confusion and frustration are important parts of the effortful problem solving needed to successfully surmount challenges and learn (D'Mello & Graesser, 2014b). However, intense or prolonged confusion and frustration can lead to anxiety and possibly despair (Zeidner, 2007). When this occurs, students are at risk of becoming disinterested and disillusioned which can lead to boredom and eventual disengagement (D'Mello & Graesser, 2012; Pekrun, Goetz, Daniels, Stupnisky, & Perry, 2010). Instead of engaging deeply in creative exploration, struggling and disengaged students exhibit problematic behaviors such as systematic guessing (Rodrigo et al., 2007) or looking for solutions rather than discovering them (e.g., Aleven, McLaren, Roll, & Koedinger, 2006; Nelson-Legall, 1987). These behaviors associated with boredom—and the negative affect that triggers them (Baker, D'Mello, Rodrigo, & Graesser, 2010)—lead to poorer learning, lower self-efficacy, diminished interest, and increased attrition (Csikszentmihalyi, 1975; Mann & Robinson, 2009; Patrick, Skinner, & Connell, 1993; Perkins & Hill, 1985; San Pedro, Baker, Bowers, & Heffernan, 2013). To prevent this negative spiral, research is needed for developing just-in-time, affect-sensitive interventions that help students persist through the "hard fun" of learning with games without losing the qualities that make games uniquely engaging and effective. An important first step in this research is to understand relationships between affect, in-game progress, and learning outcomes during game-play.

The current study is focused on modeling the relationships among affective states, in-game performance, persistence, and outcome measures of understanding physics principles, within the context of the game Physics Playground. The data we used in this paper were collected while students interacted with the game in their school's computer lab. A unique feature of the research presented here is the use of behavioral and observational measures rather than self-report measures, where possible. That is, surveys were only used for collecting student demographic information, but observational and performance-based measures were used for the other constructs. Recent reports on how the various constructs targeted in this study are currently measured underscore the overreliance on student self-report measures by the education research community (e.g., Atkins-Burnett, Fernández, Akers, Jacobson, & Smither-Wulsin, 2012; Duckworth & Yeager, 2015; Farrington et al., 2012).

Currently, the field finds itself in a loop where self-report data informs theory and program interventions, which use self-report measures to evaluate the program. However, self-report measures for constructs such as persistence (e.g., *I work hard no matter how difficult the task*) have several limitations. First, they are subject to "social desirability effects" that can lead to false reports about behavior, attitudes, and beliefs (Krosnick, 1999; Paulhaus, 1991). This refers to the tendency for people to answer in line with what society or the researchers view as favorable rather than their actual beliefs. This effect can lead to the inflation of scores related to good behaviors and/or the reduction of scores related to bad behaviors in the self-report. Another issue with self-report is that people sometimes have different conceptual understandings of the questions (e.g., what it means to "work hard" as part of a persistence question), leading to low reliability and validity (Lanyon & Goodstein, 1997). Finally, self-report items often require that individuals have explicit knowledge of their skills and dispositions (see, e.g., Schmitt, 1994), which is not always the case. People may find it difficult to accurately score themselves along the scales provided in a self-report (e.g., the ambiguity between good and very good) because they possess different levels of knowledge about themselves and/or different personalities (e.g., some are humble while others are more confident about themselves) and these need to be compared to some implicit standard, which also varies from respondent to respondent. All of these weaknesses may undermine self-report as a measure for these constructs. At the very least, they motivate a more objective approach to measurement.

In this paper, we interweave disparate sources of objective data from students as they played Physics Playground. These include: (a) log file data of in-game progress, (b) observational data on student affect during gameplay, (c) observations of on-task and off-task behaviors during game-play, (d) performance-based measures of trait persistence, (e) estimates of prior knowledge of Newtonian physics, and (f) learning outcome (i.e., physics understanding). Our main analytical approach involved the construction of structural equation modeling (SEM) to simultaneously examine the relative contribution of all of these variables to the development of conceptual physics understanding.

Structural equation modeling (SEM) is a family of related statistical techniques for testing hypothesized relationships among variables. SEM is often used as a confirmatory (a priori) technique where theory drives the specification of the model (Hoyle, 1995); but exploratory SEM also exists (e.g., Asparouhov & Muthén, 2009). Generally, SEM analysis focuses on the fit of the data to the theoretical model (Schumacker & Lomax, 1996). SEM also allows the analyst to make quantitative estimates of model parameters in addition to estimating goodness of fit. An advantage of using SEM over other statistical analyses that integrate multiple relationships into a single model is the inclusion of latent variables and the ability to extract measurement error in the analysis (Cudek, du Toit, & Sörbom, 2001).[1]

2. Method

2.1. Participants

Our sample consisted of 137 8th and 9th grade students (57 male, 80 female) who were compensated with a $25 gift card for participating in this study. Students were enrolled at a large K-12 school with a diverse population located in the southeastern U.S. We selected students in the 8th and 9th grades because of the alignment of the Physics Playground content to the Next Generation Sunshine State Standards relating to Newtonian Physics, at those grade levels. The students played Physics Playground individually on computers located in the school's lab. They played in groups of about 20 students in each of the 7 periods per day (i.e., one group of about 20 students during Period 1, another group during Period 2, and so on). There were four sessions altogether across four days. The first session, Day 1, involved completing assessments (i.e., pretest of physics, persistence performance measure, and demographic survey) and engaging in an orientation to the game. The second and third sessions, Days 2 and 3, entailed playing the game for the full approximately hour-long sessions; and the fourth session, Day 4, was divided into playing the game and then completing additional assessments (i.e., physics posttest and questionnaire about the game).

2.2. Physics Playground

Physics Playground (PP, formerly “Newton's Playground,” see Shute & Ventura, 2013) is a 2D educational video game that was developed to measure and help support middle and high school students' learning of conceptual physics related to Newton's laws of force and motion, mass, gravity, potential and kinetic energy, and conservation of momentum. A problem (or level) in PP requires the student to guide a green ball to a red balloon. The primary way to move the ball is by creating simple machines or “agents of force and motion” as they are called in the game (i.e., ramps, levers, pendulums, and springboards)—drawn with colored lines using the mouse—that “come to life” on the screen. A ramp is any line drawn that helps to guide the ball in motion (e.g., to get the ball over a hole). A lever rotates around a fixed point, usually called a fulcrum, and is useful when a student wants to move the ball vertically. A swinging pendulum directs an impulse tangent to its direction of motion, and is typically used for exerting horizontal force. A springboard stores elastic potential energy provided by a falling weight, and is useful when a student wants to move the ball vertically. Everything in the game obeys the basic rules of physics relating to gravity and Newton's laws.

Fig. 1 displays a sample problem in PP (on the left). Here, the student must draw a pendulum using a pin (little black circle) to make it swing down to hit the ball (surrounded by a heavy container hanging from a rope). In the depicted solution (on the right), the student drew a pendulum that swings down to move the ball. To succeed, the student should manipulate the mass distribution of the club and the angle from which it was dropped, to produce just the right amount of force to get the ball to the balloon.

PP includes seven playgrounds (each one containing 10–11 problems, for a total of 74 problems) that get progressively more difficult. Each problem is designed to be optimally solved by a particular agent (or agents) as discussed above. The difficulty of a problem is based on a number of factors including the relative location of ball to balloon, obstacles between the ball and balloon, and the number of agents required to solve the problem. PP also includes tutorial videos that show students

[1] See Kline (2011) for an excellent introduction to structural equation modeling, from concept to techniques.

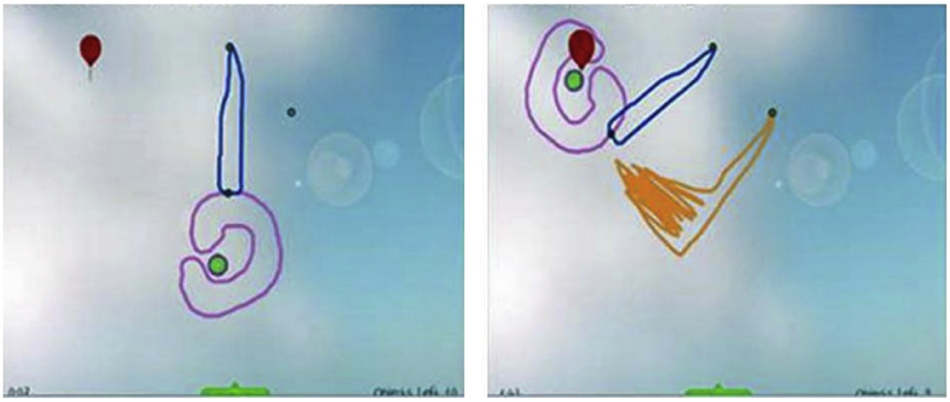

Fig. 1. Pendulum solution for a simple Physics Playground problem.

how to create and use the various agents; these videos can be watched at any time during gameplay. PP is nonlinear in that students have complete choice in selecting playgrounds and levels; however, there is a natural ordering to the interface that many students follow. Progress in the game is represented by silver and gold trophies, which are displayed in the top left part of the screen. While a silver trophy is obtained for any solution to a problem, students earn a gold trophy if a solution is under a certain number of objects (the threshold varies by problem, but is typically < 3). Achieving a gold trophy (which is worth double the points of a silver trophy) requires the efficient solution of a problem using the agent or agents that problem was designed to involve. PP maintains detailed log files that record the problem, student actions and when they occurred, system responses, trophies awarded, and so on.

2.3. Procedure

Students played Physics Playground for about 2.5 h in total (across the four sessions spanning 4 day as described earlier). The study took place in one of the school's computer labs. The computer lab contained about 30 computers, each with a monitor, mouse, keyboard, webcam, and headphones.

We administered a 15-min qualitative physics pretest during the first session and an isomorphic posttest at the end of the fourth session (both online). We also administered an online performance-based measure of persistence during the first session (see Ventura & Shute, 2013 for a validation study of the measure) as well as a demographics survey collecting information about students' age, gender, and so on. After completing the Day 1 measures, the students were introduced to Physics Playground. The instructions were as follows:

> *You will now play a video game called Physics Playground. In this game your goal is to get a ball to a balloon. In each problem you will need to draw what we call "agents" to help move the ball around the screen. Please begin with playground 1 which shows you the basics of the game. Also at any time you can click on the text at the bottom of the screen labeled "agent menu" to watch a video of each agent. When you solve a problem you will be given a gold or silver trophy for the agent you used to solve the problem. You get 1 point for a silver trophy and 2 points for gold trophy. Your total score for each agent is displayed in the top left part of the screen. Your overall goal in the game is to get as many points as possible for each agent. Good luck!*

Students began the first session by watching the agent tutorial videos and then were instructed to begin playing playground 1. After students finished playground 1 they were allowed to play any playground they wanted but were told that higher numbered playgrounds are harder. Proctors were instructed to tell students to watch the agent tutorial videos if they were stumped on a problem.

2.4. Measures

2.4.1. Physics understanding

We used a qualitative physics test consisting of 32 pictorial multiple choice items. Its purpose is to assess implicit knowledge of Newton's three laws, balance, mass, conservation and transfer of momentum, gravity, and potential and kinetic energy (see Masson, Bub, & Lalonde, 2011; Reiner, Proffit, & Salthouse, 2005). We split the qualitative physics test into two matched forms that were counterbalanced between pretest and posttest (Form A = 16 items; Form B = 16 items). For example, Fig. 2 shows an item involving a pendulum. The correct answer is "B." Reliability for the physics test was acceptable (Form A: $\alpha = .72$; Form B: $\alpha = .73$; see Shute, Ventura, & Kim, 2013). Two physics experts reviewed the tests and provided recommendations. The tests were then piloted with students prior to administration.

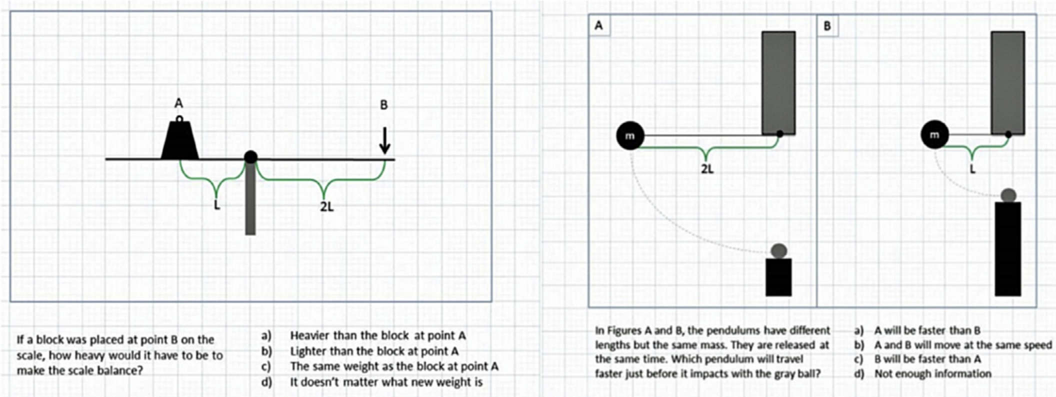

Fig. 2. Two example test items from the qualitative physics test.

2.4.2. Persistence

Following the completion of the physics measure on Day 1, we administered a short performance-based measure of persistence (PMP) that measures how much effort people exert in difficult tasks (Ventura, Shute, & Zhao, 2012). The PMP is administered online (in an Internet browser), and presents a variety of hard and easy problems (e.g., picture comparison tasks) one at a time over a series of trials. Individuals make their response and press the "guess" button. If the answer is incorrect, the screen displays "incorrect" and the individual can try again (for up to 180 s). At any time the individual can also choose to select the "skip" button to leave the current trial and go on to the next one. If the individual guesses correctly, the person is told that he or she is correct. A trial is classified as "solved" if the person accurately completes the trial. A trial is classified as unsolved if the person skips the trial or is timed out after 180 s. Time spent on unsolved trials is the critical information in the PMP that informs the assessment of persistence. Although the time spent on solved trials is likely a function of persistence as well, it may be dependent on background knowledge or ability in relation to the respective problem.

In this study, we used 6 problems—3 easy and 3 impossible (i.e., where students are told that four differences exist when there are really only three differences – see Fig. 3 for an example). The instrument has been validated previously (see Ventura & Shute, 2013) against a game-based assessment of persistence (significantly correlated) and a self-report survey of persistence (not significantly correlated).

2.4.3. Observation of affective states

Students' affective states and on-task vs. off-task behaviors were observed during their interactions with Physics Playground, using the Baker-Rodrigo-Ocumpaugh Monitoring Protocol (BROMP) for field observations (Ocumpaugh, Baker, & Rodrigo, 2015). In BROMP, trained observers perform live affect and behavior annotations by observing students one at a time, using a round-robin technique (observing one student until visible affect is detected or 20 s have elapsed and moving on to the next student). The observers use peripheral vision or side glances and make a holistic judgment of the students' affect based on facial expressions, speech, body posture, specific gestures, and information on the computer screen (e.g., whether a student is progressing or struggling). It is not always possible to observe both affect and behavior in situations where students cannot be observed (e.g., bathroom breaks, occlusions) or where the observer is not confident about an observation; in these cases the observer records a "?".

Observations of different students are conducted in a pre-determined order to maintain a representative sampling of students' affect, rather than focusing on the most interesting (but not most prevalent) things occurring in the classroom. Every BROMP observer is trained and tested with respect to the protocol and must achieve agreement of kappa ≥ 0.6 with a certified BROMP observer before being certified. The coding process is implemented using the HART application for Android devices (Ocumpaugh et al., in press), which enforces the protocol while facilitating data collection. In this study, observation-codes recorded in HART were synchronized with the videos recorded on the individual computers and with log-files of gameplay using Internet time servers.

It should be noted that there are many possible affect annotation schemes, each with their strengths and weaknesses (see Porayska-Pomsta, Mavrikis, D'Mello, Conati, & Baker, 2013 for a recent review). BROMP was selected for this study because it affords real-time affect annotation of a large number of students without interrupting or biasing the measures by asking students to self-report affect. It has been shown to achieve adequate reliability across dozens of studies with a variety of learning environments (Ocumpaugh, Baker, & Rodrigo, 2015) and from observing students during the first day of data collection. However, due to observer error, delight was not coded during the first and second day of data collection, and will not be included in this paper's analyses. In addition to affect, student behaviors were coded as *on task* when looking at their own computer and playing the game, *on-task conversation* when conversing with other students about what was happening

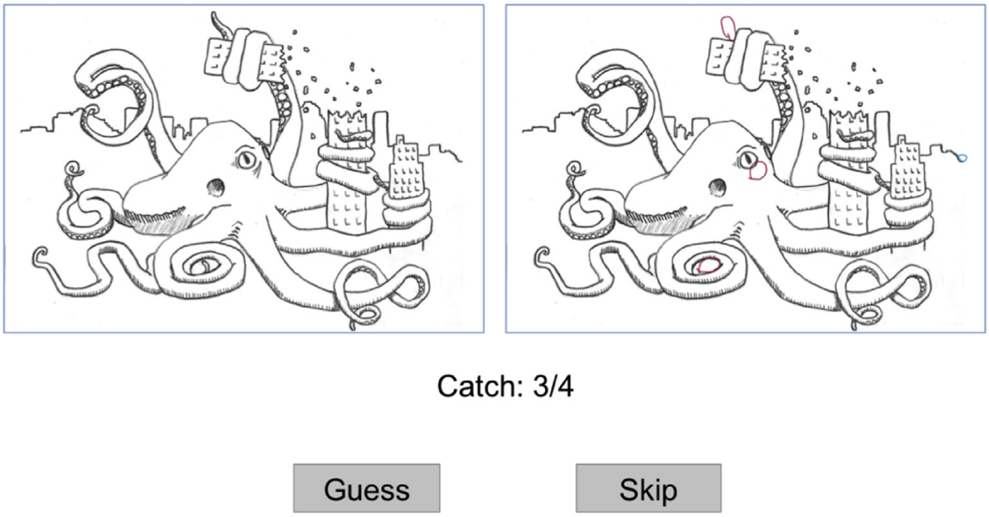

Fig. 3. Screenshot of picture-comparison item (persistence measure).

on their own or others' screens or about physics concepts, and *off task* in other situations (e.g., task-unrelated conversation, watching other students without conversation, using a cellphone).

Webcams were used to capture student facial expressions for use in further analysis. This data is beyond the scope of the current paper's research questions, but see Bosch et al. (2015) and D'Mello et al. (2015).

3. Results

We first examine whether or not the participating students learned any physics-related knowledge from playing the game. Next, we present details on the BROMP coding, before proceeding with the SEM analyses.

3.1. Learning

Did students who played Physics Playground for 2.5 h actually learn any physics principles (despite the fact that there was no instructional support in the game)? We began by examining the pretest and posttest data by form (i.e., Forms A and B, counterbalanced between pretest and posttest). The range of possible scores on each test was from 0 to 16 (the latter representing a perfect score). Because pretest scores indicated that Form B was more difficult than Form A, we equated the scores using the linear linking function (Holland, Dorans, & Petersen, 2007; Kolen & Brennan, 2004).[2] Using the equated scores, we then computed a repeated measures ANOVA on the pretest and posttest data. The ANOVA indicated a significant increase from pretest ($M = 6.99$; $SE = 0.18$; 95% CI [6.65–7.35]) to posttest ($M = 7.50$, $SE = 0.17$; 95% CI [7.20–7.83]); $F(1, 127) = 7.70$, $p = .006$. Partial eta-squared = 0.06, a medium effect size. Moreover, when we examined gender differences relative to pretest to posttest gains, the interaction was not significant: $F(1, 126) = 0.67$; $p > .05$. This suggests that both males and females improved comparably. These results replicate prior findings using the game (e.g., Shute et al., 2013).

3.2. Affective states and task-related behaviors

The affective states we observed (via BROMP observations) included: frustration, confusion, boredom, engaged concentration, and delight. We additionally captured the following behaviors: off-task behaviors, on-task gameplay, and on-task conversation about the game. We obtained 1767 observations of affective states and 1899 observations of on-task/off-task behavior during the two days of observational data used in this study (i.e., Days 2 and 3).

The overall number of observations made during gameplay, per affective state across Days 2 and 3 were: engaged concentration (1378; 78%), frustration (212; 12%), boredom (88; 5%), confusion (35; 2%), and delight (34; 2%). On-task behavior occurred 77% of the time, on-task conversation 18%, and off-task behavior was seen in just 5% of the observations, a much lower proportion than seen for other types of educational technology (e.g., Cocea, Hershkovitz, & Baker, 2009) or in traditional classrooms (Karweit & Slavin, 1982). The distributions of affective state variables were highly skewed, so for our SEM analysis we transformed the data within each affective state into five levels (1–5) based on the range of data for that

[2] The transformation involved setting z-scores of the two forms to be equal such that $(A - Mean_A)/SD_A = (B - Mean_B)/SD_B$. Therefore, the linear equating transformation from Form B to A is: $\mathbf{B} = Mean_B + [SD_B * ((A - Mean_A)/SD_A)]$.

Table 1
Correlations among affective variables.

	Confusion	Boredom	Engaged concentration
Frustration	0.29**	−0.18*	−0.76*
Confusion		0.01	−0.40**
Boredom			-0.36**

**p < .01 level; *p < .05.

particular state. Note that we did not include on-task/off-task behavior in the SEM analysis since students were almost always on-task, thus there was very little variance in the data. Delight was also excluded as it was only recorded during Day 3.

Our targeted affective variables are inter-related, as shown in the correlation matrix in Table 1. Engaged concentration was significantly and negatively related to frustration, confusion, and boredom. That is not surprising as those three affective states (i.e., frustration, confusion, and boredom) are typically considered to be negatively valenced. Second, frustration was positively related to confusion, but negatively related to boredom, complementing research that shows that confusion is frequently seen as a precursor to frustration (see D'Mello & Graesser, 2012) and these states tend to co-occur (Bosch & D'Mello, 2014). The negative relationship between frustration and boredom can be understood as follows. A person needs to be adequately invested in the learning process to experience frustration, but those who are experiencing boredom may not be sufficiently invested to even experience frustration. Finally, there was no relation between confusion and boredom.

3.3. Structural equation modeling

We conducted structural equation modeling (SEM) to predict the learning outcome from our key variables: prior knowledge (i.e., pretest score), persistence, affective states, and in-game performance data. Our approach started out as confirmatory, but rather than leaving inappropriate links in a theoretical model, we wanted to test and refine our theoretical model. Consequently, we used an approach that was primarily confirmatory but also took into consideration which of several principled models was most appropriate. Our SEM models were estimated using lavaan—an R-package for structural equation modeling—using maximum likelihood robust (MLR) estimation to fit the non-normal data to our hypothesized model (Rosseel, 2012). Full Information Maximum Likelihood (FIML) estimation was used to handle the occasional missing values.

Prior to conducting the full SEM, we conducted confirmatory factor analysis (CFA) to estimate our in-game and learning constructs, each construct having four indicators. The results of the CFA yielded the following: χ^2 (19) = 17.8, p = .54. The Comparative Fit Index (CFI) = 1.00, the root mean square error of approximation (RMSEA) = 0.00 (90% CI: 0.00–0.069), and the standardized root mean residual (SRMR) = 0.04, all of which suggest a good fit of the two single-factor constructs (i.e., in-game performance and learning) with the data.[3]

Having verified the adequacy of the measurement model, we proceeded with the structural model by specifying directionality among our two factors (i.e., in-game performance and learning outcome) and six observed variables (i.e., pretest, persistence, frustration, engaged concentration, confusion, and boredom). The rationale for the hypothesized model was that individuals' level of persistence (as measured via the picture comparison task) and incoming knowledge (physics pretest score) would influence their affective states as well as how they played the game—such as fueling engaged concentration that would enable a person to stick with a hard level and try various ways to solve it. Once immersed in gameplay, affective states would change dependent on difficulties and successes in solving the games' levels. We hypothesized that affective states would influence performance in the game as evidenced by the in-game construct, which consisted of the number of gold trophies received for ramp, lever, pendulum, and springboard solutions. Covariance relationships were added between the affect variables since they are expected to covary. Finally, we predicted that the degree to which an individual was successful in the game (gold trophy data) would positively predict learning outcome (i.e., posttest score).

We tested three similar models based on the rationale above. The first two models were the same except for the relationships involving affective states. In Model 1, we linked the four affective states to in-game behaviors and also directly to the outcome measure. In Model 2, affective states were only linked to the in-game measure. We did not include the direct relationships among the affective states and learning outcome because the results from the first model showed no significant paths. Findings revealed that Model 2 was not statistically different from Model 1 (χ^2 (4) = 2.38, p = .67). Model 2 was a simpler model, which may be preferred.

[3] Structural Equation Modeling (SEM), CFA specifically, relies on several statistical tests to determine the adequacy of model fit to the data. The chi-square test indicates the amount of difference between expected and observed covariance matrices. A chi-square value close to zero indicates little difference between the expected and observed covariance matrices. The Comparative Fit Index (CFI) compares the fit of the baseline model (which assumes there is no causal relationship among variables) to the theoretical model. CFI ranges from 0 to 1 with a larger value indicating better model fit. Acceptable model fit is indicated by a CFI value of 0.90 or greater (Hu & Bentler, 1999). Root Mean Square Error of Approximation (RMSEA) is equal to the discrepancy function adjusted for model degrees of freedom. A smaller RMSEA value indicates better model fit. Acceptable model fit is indicated by an RMSEA value of 0.06 or less (Hu & Bentler, 1999; Kline, 2011).

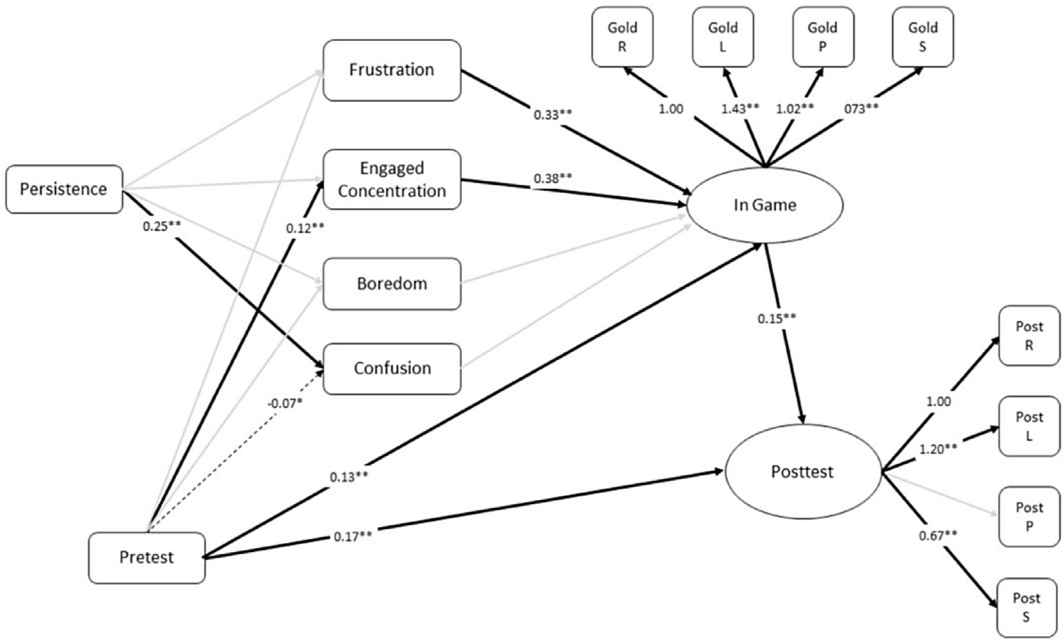

Fig. 4. Model 2 with parameter estimates. Notes: (1) $^{**}p < .05$; $^{*}p < .10$; (2) Black links = positive coefficients, dashed links = negative coefficients, grey links = non-significant; (3) All variables are positively scaled; (4) R = ramp; L = lever; P = pendulum; S = springboard.

Model 2 included two latent measurement models for the in-game measure and learning, and six observed variables. Each of the latent constructs had four indicators measuring the construct. The fit indices of Model 2 were: χ^2 (61) = 57.6, $p = .60$; CFI = 1.00; RMSEA = 0.00 (90% confidence interval: 0.00–0.05), and SRMR = 0.05. Fig. 4 shows the results of the parameter estimates from the SEM analysis.

According to our hypothesized model, learning—as measured via the physics posttest—is a function of incoming knowledge (pretest) and in-game performance, while in-game performance is a function of incoming knowledge and various affective states. The learning construct for our measurement model, labeled "posttest" in Fig. 4, shows that the lever and springboard subscales are significantly related to learning (the ramp sub-scale was fixed at 1.0 to identify the model). For the in-game performance measure, the factor loadings for gold trophies related to lever, pendulum, and springboard were significant (the ramp gold trophy was set to 1.0).

We speculated that persistence would differentially influence various affective states (e.g., positively predicting engaged concentration and frustration), but that was not the case. Thus, we created our final model (Model 3), shown in Fig. 5, which eliminated persistence and additionally eliminated one of the indicators of our learning construct (i.e., pendulum posttest[4]) given its non-significant relation to learning. The fit indices of the revised measurement model were good χ^2 (13) = 12.8, $p = .46$; CFI = 1.00; RMSEA = 0.00 (90% confidence interval: 0.00–0.08), and SRMR = 0.04, so we proceeded with the structural model (Model 3).

Fit indices for Model 3 were good: χ^2 (42) = 43.2, $p = .42$; CFI = 0.99; RMSEA = 0.02 (90% confidence interval: 0.00–0.06), and SRMR = 0.04. Within the model, pretest was a positive, significant predictor of in-game performance ($b = 0.13$, $p = .003$). Additionally, frustration ($b = 0.33$, $p = .048$) and engaged concentration ($b = 0.38$, $p = .012$) positively predicted in-game performance. Pretest ($b = 0.17$, $p < .001$) and the in-game measure ($b = 0.15$, $p = .031$) were both significant predictors of learning, as expected. Pretest also significantly predicted engaged concentration ($b = 0.12$, $p = .035$).

Several indirect paths were also tested: *Engagement → In-Game Performance → Posttest* ($p = .042$) and *Frustration → In-Game Performance → Posttest* ($p = .086$). These suggest that there may be two paths to success—one through engaged concentration and the other through frustration (albeit the latter is marginally significant). Two additional indirect paths were tested: *Pretest → In-Game Performance → Posttest* ($p = .045$); and one path that was not significant: *Pretest → Engagement → In-Game Performance → Posttest* ($p = .147$). In short, affective states influence learning via shaping in-game performance while incoming knowledge influences learning directly as well as indirectly via in-game performance.

[4] Results were identical when we tested a version of Model 3 with persistence eliminated but with pendulum posttest included as an indicator of our learning construct.

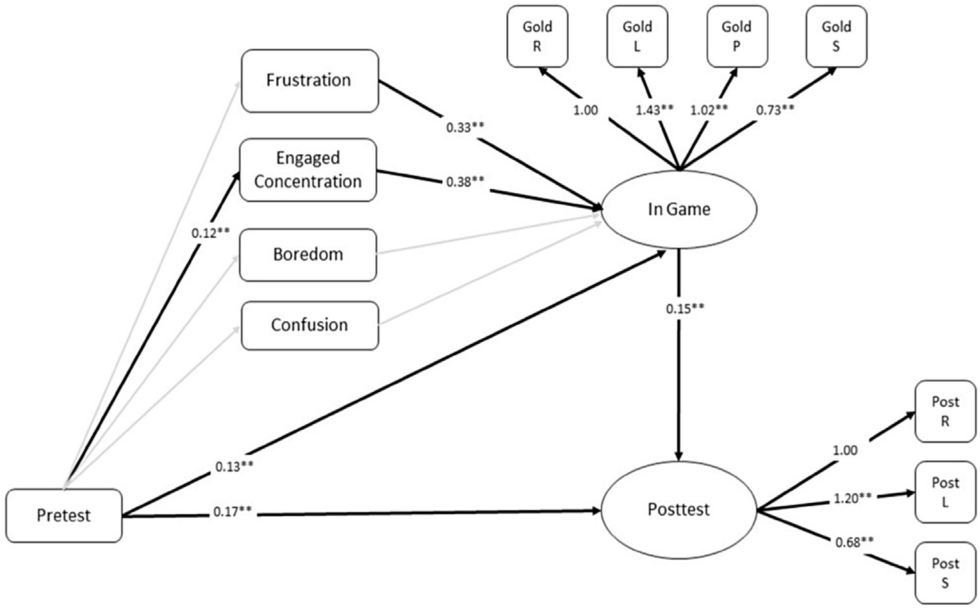

Fig. 5. Model 3 with parameter estimates. Notes: (1) **$p < .05$; *$p < .10$; (2) Black links = positive coefficients, dashed links = negative coefficients, grey links = non-significant; (3) All variables are positively scaled; (4) R = ramp; L = lever; P = pendulum; S = springboard.

4. Discussion

Well-designed games are intentionally created to support motivation and engagement, but the mechanisms by which these constructs engender learning are not entirely understood. In particular, researchers do not fully understand how these constructs interact with students' prior knowledge and performance within the system. In this research, we sought to explore the various relationships among incoming knowledge, persistence, affective states, in-game progress, and learning.

In terms of learning from the game, we found that students did significantly improve from pretest to posttest, and the 95% confidence intervals show clear differences between the pretest and posttest score distributions. Thus an improvement from 6.99 to 7.50 may not seem like much, but it occurred without any instructional intervention. In another study (Kim & Shute, in press) using undergraduate students rather than middle school students, the test means were slightly higher than reported in the current study (pretest = 10.4; posttest = 11.0), but the scores are still far from perfect, likely due to the conceptual difficulty of the content. Future research in this area can focus on designing and incorporating instructional interventions into the game (i.e., visualizations, explanations) towards the goal of personalized and more formal physics learning.

The results of our SEM analyses revealed that incoming knowledge predicted engaged concentration, in-game progress (i.e., receipt of gold trophies for ramp, lever, pendulum, and springboard), and learning outcome. Perhaps more interesting is the finding that the affective states of frustration and engaged concentration influenced the learning outcome indirectly by way of in-game performance. We did not find that either confusion or boredom predicted in-game progress, likely due to their states being rarely observed among the students in this study, thus having no meaningful variability. The findings for engaged concentration and frustration confirm our basic hypothesis.

The influence of engaged concentration on in-game behaviors and learning might be expected, but the effects of frustration are a little more interesting. Educators often work to minimize student frustration, but a number of studies suggest that this tendency may have unintended consequences. In work with educational games, Gee (2007) has suggested that adaptive challenges could help players to achieve a state of "pleasurable frustration," which motivates players to overcome obstacles, constraints, and hard problems. Players who experience this pleasurable frustration may persevere because reaching for goals and ultimately succeeding is highly rewarding. McGonigal (2011) has referred to this as a positive kind of stress, called eustress, which is actually good for us, providing a sense of motivation and desire to succeed. We see this pleasant frustration (or eustress) as a positive aspect of video games because it shows that students are being pushed to their limits, a requirement for teaching in the zone of proximal development (Vygotsky, 1978). Some degree of repeated frustration can also prepare students for tolerating frustration, a common emotional response in education and in life in general.

The challenge for researchers and designers is to determine the appropriate level of frustration for different types of learners. Given that one can navigate through playgrounds with varying levels of difficulty, it is highly likely that students will select different problem sequences. Students who want to be challenged may select higher numbered playgrounds, while risk-averse students may avoid choosing more challenging levels altogether. Future research should investigate the potential relationship between affect and student personalities, in order to better support how different types of learners deal with frustration.

The immediate goal of the research was to examine the direct and mediating variables influencing learning. Once identified, these can be used as the basis for instructional support. For instance, if a student is experiencing too many consecutive periods of frustration, or has transitioned from frustration to boredom, some type of motivational support or intervention may be necessary to get the student to reengage. There are a variety of potential responses a game can take when a student encounters frustration, from supportive messages (D'Mello, Craig, Fike, & Graesser, 2009; Klein, Moon, & Picard, 2002) to temporarily switching the student to a less challenging activity. Similarly, interventions might be designed around the finding that confusion, which did not lead to learning in these models, was more common among those students who had shown high levels of persistence. This relationship can be seen in Fig. 4. Prior research has shown that confusion appears to be important in some learning contexts (D'Mello & Graesser, 2012). This relationship suggests that students might benefit from targeted interventions, based on threshold levels of confusion, to prevent wheel-spinning (Beck & Gong, 2013) which occurs when students are unable to successfully resolve their confusion (D'Mello & Graesser, 2014b). These interventions might be designed to help students acquire the metacognitive awareness needed to differentiate between productive and unproductive struggling so that they can make more efficient use of their learning opportunities. Future research in this area may involve analyzing particular sequences of affective states relative outcome(s) rather than aggregated affective data as we used. This could lead to more particular and effective rules to drive affective and other types of learning support.

Although predicted, we did not find any meaningful relationship involving persistence and other variables (e.g., affective states) in the SEM models. This finding may have been due to a number of factors. First, in this study we used a shortened version of the performance-based measure of persistence (i.e., 3 impossible and 3 easy items rather than 7 impossible and 7 easy items as we have used in past experiments; see Ventura & Shute, 2013) which may have reduced the reliability of the measure. Second, we used aggregated percentages of affective states which could have further clouded any relations. Third, while we found that our persistence measure significantly related to particular gameplay variables in a related study (Shute et al., 2013), there was no correlation between persistence and any gameplay variable in the current study, casting more doubt about the quality of the shortened persistence measure. Finally, the lack of relations between persistence and other variables in our study may have been caused by the linear-relations constraint in SEM, discussed next.

This study is not without limitations. One limitation of SEM is that the techniques examine first-order (linear) relationships between variables (albeit, power transformations may be made if a relationship between two variables seems quadratic). Some of our data—such as the persistence variable—may have actually had quadratic relations with other variables. For instance, too little and too much persistence may be sub-optimal for success in the game and ultimately learning, while some intermediate amount may be best. We focused on linear relationships as an initial step in this research. An important next step would be to consider nonlinear relationships among constructs as well.

A second limitation pertains to the use of SEMs in general. Specifically, we proposed a structural model informed by theory and showed that it had adequate fit from a statistical perspective. However, it is possible that alternate structural models would also yield similarly good fits on the same data. Therefore, our proposed model and results should be taken as one possible interpretation of the data. Other interpretations are also possible, akin to multiple theories being developed to explain some phenomenon. An important future work item would be to structurally specify alternate models and compare their fit to the present data.

The sample size of 137 students might also be a limitation as SEMs generally require large amounts of data due to the number of parameter estimations needed (Kline, 2011). One common rule-of-thumb when using SEM is to use sample sizes of at least 200 (or about 5–10 cases per parameter; see Kline, 2011). However, more recent papers suggest that smaller sample sizes are sufficient (i.e., Sideridis, Simos, Papanicolaou, & Fletcher, 2014; Wolf, Harrington, Clark, & Miller, 2013).

In summary, by understanding the ways that affect interacts with student knowledge, persistence, performance, and learning, we can learn to design systems that not only scaffold affect in order to promote better learning, but help to create more persistent and more successful learners. In doing so, we can help learners to develop not just their knowledge of physics and other content areas, but the types of self-regulated learning skills that will help them throughout their lives.

Acknowledgments

We'd like to thank the Bill & Melinda Gates Foundation (SOL1071343/APP181499) for their generous support provided for this research, particularly our program officer Ed Dieterle. We also would like to acknowledge the awesome support of our statistics colleague, Yanyun Yang, and graduate students—Lubin Wang and Weinan Zhao, as well as the teachers and staff at FSUS, and the students who participated in the project.

References

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: a model of help-seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*, 101–128.

Andres, J. M. L., Rodrigo, M. M., Sugay, J. O., Baker, R. S., Paquette, L., Shute, V. J., et al. (2014). An exploratory analysis of confusion among students using Newton's Playground. In C. C. Liu, et al. (Eds.), *Proceedings of the 22nd International Conference on Computers in Education* (pp. 65–70). Japan: Asia-Pacific Society for Computers in Education.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397–438.

Atkins-Burnett, S. A., Fernández, C., Akers, L., Jacobson, J., & Smither-Wulsin, C. (2012). *Landscape analysis of non-cognitive measures*. Princeton, NJ: Mathematica Policy Research.

Baker, R., D'Mello, S., Rodrigo, S., & Graesser, A. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223–241.
Beck, J. E., & Gong, Y. (2013). Wheel-spinning: students who fail to master a skill. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education* (pp. 431–440). Berlin/Heidelberg: Springer.
Bosch, N., & D'Mello, S. K. (2014). It takes two: momentary co-occurrence of affective states during computerized learning. In S. Trausan-Matu, K. Boyer, M. Crosby, & K. Panourgia (Eds.), *Proceedings of the 12th International Conference on Intelligent Tutoring Systems (ITS 2014)* (pp. 638–639). Switzerland: Springer International Publishing.
Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., et al. (2015). Automatic detection of learning-centered affective states in the wild. In *Proceedings of the 2015 International Conference on Intelligent User Interfaces*. New York, NY: ACM.
Calvo, R. A., & D'Mello, S. K. (Eds.). (2011). *New perspectives on affect and learning technologies*. New York: Springer.
Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. (2014). *Digital games, design, and learning: A systematic review and meta-analysis*. Menlo Park, CA: SRI International.
Cocea, M., Hershkovitz, A., & Baker, R. S. J. D. (2009). The impact of off-task and gaming behaviors on learning: Immediate or aggregate?. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education, Brighton, UK* (pp. 507–514).
Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Journal of Applied Artificial Intelligence, 16*(7–8), 555–575.
Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction, 19*(3), 267–303.
Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.
Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper & Row.
Cudek, R., du Toit, S., & Sörbom, D. (Eds.). (2001). *Structural equation modeling: Present and future. Lincolnwood, IL*. Scientific Software International.
D'Mello, S. K., Bosch, N., Baker, R., Ocumpaugh, J., Shute, V. J., Ventura, M., et al. (2015, April). Facial features for automatic detection of student affect with Newton's Playground in the wild. In *Paper presented at AERA 2015, Chicago, IL*.
D'Mello, S. K., Craig, S. D., Fike, K., & Graesser, A. C. (2009). Responding to learners' cognitive-affective states with supportive and shakeup dialogues. In J. A. Jacko (Ed.), *Human-computer interaction. Ambient, ubiquitous and intelligent interaction* (pp. 595–604). Berlin/Heidelberg: Springer.
D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145–157.
D'Mello, S. K., & Graesser, A. C. (2014a). Confusion. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 289–310). New York, NY: Routledge.
D'Mello, S., & Graesser, A. C. (2014b). Inducing and tracking confusion and cognitive disequilibrium with breakdown scenarios. *Acta Psychologica, 51*, 106–116.
Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher, 44*(4), 237–251.
Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.
Eisenberger, R. (1992). Learned industriousness. *Psychological Review, 99*, 248–267.
Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., et al. (2012). *Teaching adolescents to become learners. The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago, IL: University of Chicago Consortium on Chicago School Research.
Fiedler, K., & Beier, S. (2014). Affect and cognitive processes in educational contexts. In R. Pekrun, & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 36–56). New York, NY: Routledge.
Gee, J. P. (2007). *Good video games + good learning: Collected essays on video games, learning and literacy*. New York, NY: Peter Lang.
Holland, P. W., Dorans, N. J., & Petersen, N. S. (2007). Equating test scores. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of statistics: Psychometrics* (pp. 169–203). Amsterdam, The Netherlands: Elsevier.
Hoyle, R. H. (Ed.). (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks: Sage Publications.
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies, 65*(8), 724–736.
Karweit, N., & Slavin, R. E. (1982). Time-on-task: Issues of timing, sampling, and definition. *Journal of Educational Psychology, 74*(6), 844–851.
Kim, C., & Pekrun, R. (2014). Emotions and motivation in learning and performance. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 65–75). New York, NY: Springer.
Kim, Y. J., & Shute, V. J. (2015). The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment (in press) *Computers & Education, 87*, 340–356.
Klein, J., Moon, Y., & Picard, R. (2002). This computer responds to user frustration – theory, design, and results. *Interacting with Computers, 14*(2), 119–140.
Kline, R. B. (2011). *Principles and practice of structural equation modelling* (3rd ed.). New York, NY: Guilford.
Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking. Methods and practices* (2nd ed.). New York, NY: Spring-Verlag.
Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*(1), 537–567.
Lanyon, R. I., & Goodstein, L. D. (1997). *Personality assessment* (3rd ed.). New York, NY: Wiley.
Mann, S., & Robinson, A. (2009). Boredom in the lecture theatre: an investigation into the contributors, moderators and outcomes of boredom amongst university students. *British Educational Research Journal, 35*(2), 243–258.
Masson, M. E. J., Bub, D. N., & Lalonde, C. E. (2011). Video-game training and naive reasoning about object motion. *Applied Cognitive Psychology, 25*, 166–173.
McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. New York: Penguin Press.
Nelson-Legall, S. A. (1987). Necessary and unnecessary help-seeking in children. *Journal of Genetic Psychology, 148*(1), 53–62.
Ocumpaugh, J., Baker, R. S., & Rodrigo, M. M. T. (2015). *Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 technical and training manual*. Technical Report. New York, NY: Teachers College, Columbia University. Manila, Philippines: Ateneo Laboratory for the Learning Sciences.
Ocumpaugh, J., Baker, R. S., Rodrigo, M. M. T., Salvi, A., van Velsen, M., Aghababyan, A., et al. (2015). *HART: The Human Affect Recording Tool. ACM Special Interest Group on the Design of Communication (SIGDOC)*. New York, NY: ACM.
Park, O., & Lee, J. (2004). Adaptive instructional systems. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 651–685). Mahwah, NJ: Lawrence Erlbaum.
Patrick, B., Skinner, E., & Connell, J. (1993). What motivates children's behavior and emotion—Joint effects of perceived control and autonomy in the academic domain. *Journal of Personality and Social Psychology, 65*(4), 781–791.
Paulhaus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (Vol. 1, pp. 17–59). San Diego, CA: Academic Press.
Pekrun, R., Goetz, T., Daniels, L., Stupnisky, R. H., & Perry, R. (2010). Boredom in achievement settings: exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 102*(3), 531–549.
Perkins, R. E., & Hill, A. B. (1985). Cognitive and affective aspects of boredom. *British Journal of Psychology, 76*, 221–234.
Porayska-Pomsta, K., Mavrikis, M., D'Mello, S. K., Conati, C., & Baker, R. (2013). Knowledge elicitation methods for affect modelling in education. *International Journal of Artificial Intelligence in Education, 22*, 107–140.
Reiner, C., Proffit, D. R., & Salthouse, T. (2005). A psychometric approach to intuitive physics. *Psychonomic Bulletin and Review, 12*, 740–745.
Rodrigo, M. M. T., Baker, R. S. J. D., Lagud, M. C. V., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., et al. (2007). Affect and usage choices in simulation problem solving environments. *Proceedings of Artificial Intelligence in Education*, 145–152.
Rosseel, Y. (2012). lavaan: an R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

San Pedro, M. O. Z., Baker, R. S. J. D., Bowers, A. J., & Heffernan, N. T. (2013). Predicting college enrollment from student interaction with an Intelligent Tutoring System in middle school. In *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 177–184).

Schmitt, N. (1994). Method bias: the importance of theory and measurement. *Journal of Organizational Behavior, 15*, 393–398.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.

Shute, V. J., Lajoie, S. P., & Gluck, K. A. (2000). Individualized and group approaches to training. In S. Tobias, & J. D. Fletcher (Eds.), *Training and retraining: A handbook for business, industry, government, and the military* (pp. 171–207). New York: Macmillan.

Shute, V. J., Rieber, L., & Van Eck, R. (2011). Games … and … learning. In R. Reiser, & J. Dempsey (Eds.), *Trends and issues in instructional design and technology* (3rd ed., pp. 321–332). Upper Saddle River, NJ: Pearson Education, Inc.

Shute, V. J., & Ventura, M. (2013). *Measuring and supporting learning in games: Stealth assessment*. Cambridge, MA: The MIT Press.

Shute, V. J., Ventura, M., & Ke, F. (2015). The power of play: the effects of Portal 2 and lumosity on cognitive and noncognitive skills. *Computers & Education, 80*, 58–67.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of informal physics in Newton's Playground. *The Journal of Educational Research, 106*, 423–430.

Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach, & A. Lesgold (Eds.), *Adaptive technologies for training and education* (pp. 7–27). New York, NY: Cambridge University Press.

Sideridis, G., Simos, P., Papanicolaou, A., & Fletcher, J. (2014). Using structural equation modeling to assess functional connectivity in the brain: power and sample size considerations. *Educational and Psychological Measurement, 74*(5), 733–758.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Sternberg, & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3–37). New York: Cambridge University Press.

Ventura, M., & Shute, V. J. (2013). The validity of a game-based assessment of persistence. *Computers and Human Behavior, 29*, 2568–2572.

Ventura, M., Shute, V., & Zhao, W. (2012). The relationship between video game use and a performance-based measure of persistence. *Computers & Education, 60*, 52–58.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models an evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6), 913–934.

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: when students believe that personal characteristics can be developed. *Educational Psychologist, 47*(4), 302–314.

Zeidner, M. (2007). Test anxiety: conceptions, findings, conclusions. In P. Schutz, & R. Pekrun (Eds.), *Emotion in education* (pp. 165–184). San Diego, CA: Academic Press.