Computers in Human Behavior 29 (2013) 2568–2572

Contents lists available at SciVerse ScienceDirect

Computers in Human Behavior

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

The validity of a game-based assessment of persistence

Matthew Ventura *, Valerie Shute

Florida State University, 600 W College Ave, Tallahassee, FL 32304, USA

ARTICLE INFO

Article history:

Keywords:
Game-based assessment
Persistence
Qualitative physics
Formative assessment

ABSTRACT

In this study, 154 students individually played a challenging physics video game for roughly 4 h. Based on time data for both solved and unsolved problems derived from log files, we created a game-based assessment of persistence that was validated against an existing measure of persistence. We found that the game-based assessment of persistence predicted learning of qualitative physics after controlling for gender, video game experience, pretest knowledge and enjoyment of the game. These findings support the implementation of a real-time formative assessment of persistence to be used to dynamically change gameplay.

1. Introduction

There is growing evidence of video games and simulations supporting learning (e.g., Coller & Scott, 2009; Tobias & Fletcher, 2011; for a review see Wilson et al., 2009). An additional advantage of using video games and simulations in education is the vast amount of data that can be used for assessment purposes (Dede, 2005; DiCerbo & Behrens, 2012; Quellmalz, Timms, Silberglitt, & Buckley 2012; Shute & Ventura, 2013). Formative assessments embedded within a video game can enable us to more accurately provide feedback and change gameplay to maximize learning according to the ability level of the player.

In this paper, we focus on a game-based assessment for persistence, a facet of conscientiousness. Over the past 20 years or so, conscientiousness has emerged as one of the most important personality traits in predicting academic performance (e.g., Poropat, 2009) as well as in various life outcomes (e.g., Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). Persistence (i.e., industriousness in Roberts, Chernyshenko, Stark, & Goldberg, 2005; achievement in Perry, Hunter, Witt, & Harris, 2010) is a facet of conscientiousness that reflects a dispositional need complete difficult tasks (McClelland, 1961), and the desire to exhibit high standards of performance in the face of frustration (Dudley, Orvis, Lebiecki, & Cortina, 2006). Perry et al. (2010) suggest that persistence may drive the predictive validity of conscientiousness and is the facet that consistently predicts a variety of outcomes (Dudley et al., 2006; Perry et al., 2010; Roberts et al., 2005) over other facets of conscientiousness.

Persistence can play an important role in learning in a video game due to the principle of challenge (Pausch, Gold, Skelly, & Thiel, 1994). That is, challenge entails adjusting the optimal level of difficulty for a player and is consistent with the theory of the zone of proximal development (Vygotsky, 1978) which states that learning takes place right at the outer edges of one's abilities. The principle of challenge is pervasively used in video games and has been shown to engage attention and enhance learning (Lepper & Malone, 1987; Rieber, 1996; Sweetser & Wyeth, 2005). Thus video games can require persistence due to the design of progressive difficulty. This repeated exposure to challenge can positively affect persistence requiring a willingness to work hard despite repeated failure (for a review see Eisenberger, 1992; Ventura, Shute, & Zhao, 2012). For example, Eisenberger and Leonard (1980) showed that exposure to difficult tasks can improve persistence. Participants were randomly assigned to solve impossible, hard, or easy anagrams and then take the perceptual comparison task. Then participants were asked to detect as many differences as possible between two pictures. Participants in the impossible anagram condition spent the most time on the perceptual comparison task, followed by those in the hard anagram condition, and then those in the easy anagram condition. This provides evidence that exposure to difficult tasks can affect subsequent effort. The next section introduces a video game we developed that requires persistence due to its difficulty.

1.1. Qualitative physics in Newton's playground

Research into what's called "folk" physics demonstrates that many adults hold erroneous views about basic physical principles that govern the motions of objects in the world, a world in which people act and behave quite successfully (Reiner, Proffit, & Salthouse, 2005). The prevalence of these systematic errors has led some investigators to propose that incorrect performance on these tasks is due to specific "naive" beliefs, rather than to a

* Corresponding author. Address: College of Education, 3205A Stone Building, 1114 West Call Street, Tallahassee, FL 32306-4453, USA. Tel.: +1 609 851 0767.
E-mail addresses: mventura@fsu.edu (M. Ventura), vshute@fsu.edu (V. Shute).

0747-5632/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chb.2013.06.033

general inability to reason about mechanical systems (McCloskey & Kohl, 1983). Recognition of the problem has led to interest in the mechanisms by which physics students make the transition from folk physics to more formal physics understanding (diSessa, 1982) and to the possibility of using video games to assist in the learning process (Masson, Bub, & Lalonde, 2011; White, 1994).

One way to help remove misconceptions in physics is to illustrate physics principles with physical machines (Hewitt, 2009). In physics, a machine refers to a device that is designed to either change the magnitude or the direction of a force. Teaching about simple machines (e.g., lever, pulley, and wedge) is widely used as a method to introduce physics concepts (Hewitt, 2009). Research on science education also indicates that learners' hands-on experience with such machines (both virtually and physically) support applicable understanding of important physics concepts (Hake, 1998).

We developed a PC video game called Newton's Playground to help middle school students understand Qualitative Physics (Ploetzner & VanLehn, 1997). Qualitative physics as a nonverbal understanding of Newton's three laws, balance, mass, gravity, conservation of momentum, potential and kinetic energy. Newton's Playground is a 2D sandbox video game (i.e., a game design feature where the player can create objects) that requires the player to guide a green ball to a red balloon (inspired by the game Crayon Physics Deluxe). The player can nudge the ball to the left and right (if the surface is flat) but the primary way to move the ball is by drawing/creating simple machines on the screen that "come to life" once the object is drawn. Everything obeys the basic rules of physics relating to gravity and Newton's three laws of motion.

The 74 problems in Newton's Playground (NP) require the player to draw/create four simple machines: inclined plane/ramps, pendulums, levers, and springboards. All solutions are drawn with colored lines using the mouse. A ramp is any line drawn that helps to guide a ball in motion. A ramp is useful when a ball must travel over a hole. A lever rotates around a fixed point usually called a fulcrum or pivot point. Levers are useful when a player wants to move the ball vertically. A swinging pendulum directs an impulse tangent to its direction of motion. The pendulum is useful when the player wants to exert a horizontal force. A springboard (or diving board) stores elastic potential energy provided by a falling weight. Springboards are useful when the player wants to move the ball vertically. Fig. 1 displays a problem in NP. In this problem the player must draw a pendulum on a pin (i.e., 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, the player drew a pendulum that will swing down to move the ball. To succeed, the player should manipulate the mass distribution of the club and the angle from which it was dropped to accomplish just the right amount of force to get the ball to the balloon.

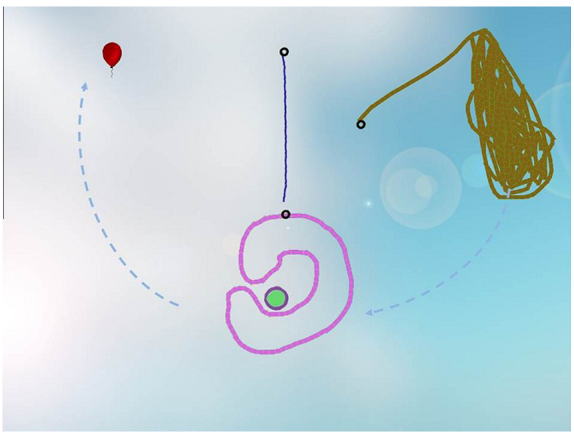

Fig. 1. Example problem in NP.

1.1.1. Other gameplay features

NP consists of 7 playgrounds (each one containing 10–11 problems) that progressively get more difficult. Each problem is designed to elicit a particular simple machine (in the game we refer to them as "agents"). The difficulty of a problem is based on a number of factors including: relative location of ball to balloon, obstacles, number of agents required to solve the problem, and novelty of the problem. NP also includes tutorial videos that show the player how to create and use the various agents. During gameplay, students have the option to watch agent tutorial videos at any time.

NP displays silver and gold trophies in the top left part of the screen which represent progress in the game. A silver trophy is obtained for any solution to a problem. Players can also receive a gold trophy if a solution is under a certain number of objects (the threshold varies by problem, but is typically <3). A player can receive one silver and one gold trophy per problem.

1.1.2. NP session logs

NP automatically uploads log files to a server for each gaming session (i.e., log activity between login and logout). The code below displays what a session log looks like for one *event* of a problem. An event collects data for a particular visit to a problem. A player may revisit a problem multiple times thus logging multiple events. Fig. 2 displays a snapshot of the NP session event log. As can be seen the session event log reports several features of gameplay in a problem. For example, "game_time" reports the total time spent on this particular visit to the problem. "Silver" reports if a silver trophy was achieved in this visit to the problem.

2. Theory

The aim of this study is to describe how we used the log file data to develop a game-based assessment of persistence. We focus on two pieces of data from the log files to inform our game-based assessment of persistence (GAP): *unsolved times* and *solved times* on problems. That is, longer times spent on difficult problems (whether they were solved or not) should indicate greater persistence (Eisenberger & Leonard, 1980; Ventura et al., 2012).

We predict that the GAP will be positively correlated to another persistence measure called the Performance Measure of Persistence (Ventura et al., 2012). The PMP measures how much time participants spend on impossible problems. The Performance Measure of Persistence (PMP) is administered on a computer and presents impossible and easy problems one at a time over a series of trials. Additionally, we expect the PMP (external to the game)


```
"time_stamp" : 12.163,
"level_path" : ".\\levels\\p4\\diving board.level",
"game_time" : 130.526001,
"pause_time" : 1.54,
"restart_count" : 7,
"object_count" : 14,
"object_limit_count" : 2,
"nudge_count" : 42,
"erase_count" : 13,
"pin_count" : 9,
"agent_vector" :"61.78 SB, 98.08 SB, 131.60 SB",
"ball_trajectory" : "<0.733, 0.427> <0.766, 0.394>….
"silver" : true,
"gold" : false,
"solved" : true
```


Fig. 2. NP session event log.

and GAP (in-game measure) to relate to a traditional self-report measure of persistence.

Like traditional measures of persistence, the GAP should also relate to achievement after controlling for other variables. Thus we predict that the GAP will positively relate to learning physics in NP after controlling for gender, video game experience, physics pretest knowledge and enjoyment. The reason we control for these variables is because we want to know the incremental validity of persistence after controlling for other variables that may affect learning.

Finally we expect that the GAP will be a more valid for players who are struggling in the game. Struggling players are constantly being challenged, which is a core requirement for eliciting persistence (Ventura et al., 2012). High performers are not being challenged as much as low performers and thus may not need to be as persistent during gameplay. We chose to use gold trophies as a measure of performance in NP since gold trophies indicate the player fully understands the simple machine needed to solve a particular problem.

3. Method

3.1. Sample and procedure

Our sample consisted of 154 8th and 9th grade students (72 male, 82 female). Students were paid $25 for participation and were tested in groups of around 20 students per session. The students played NP for around 4 h (split into five 45-min sessions across two weeks) in a large computer lab. Students were not allowed to talk to other students or to look at other students' computer screens.

We administered a qualitative physics pretest, a self-report persistence questionnaire, and a video game use question at the beginning of the study, and a physics posttest as well as the PMP at the end of gameplay (all online). After completing the pretest, the students were told about NP and that the person with the most gold trophies at the end of the study would receive a special prize (an extra $25). Students began the first session by watching the simple machine tutorial videos and then were instructed to begin playing playground 1. After participants finished playground 1, they were instructed to play any playground they wanted but were told that higher numbered playgrounds are harder. Proctors were instructed to tell players to watch the agent tutorial videos again if they were stumped on a problem.

3.2. Measures

3.2.1. Performance measure of persistence (Ventura et al., 2012)

The PMP measures how much time participants spend on particularly difficult problems. The PMP is administered online (in an Internet browser) and presents 7 impossible and 7 easy problems one at a time over a series of trials. There are two types of tasks in the PMP: anagram and picture comparison tasks. The picture comparison task is a new PMP and requires participants to find differences between two adjacent pictures. Four anagrams are impossible (i.e., anagrams that do not actually make a word) and three picture comparison items are impossible (i.e., where subject are told that four differences exist when there are really only three differences). For anagram items, participants type in their response for the anagram into a text box and press the "guess" button. For picture comparison items, participants click on differences in the pictures and click the guess button. If the answer is wrong, the screen displays "incorrect" and the individual can try again (for up to 120 s). At any time the individual can also choose to select the "skip" button to leave the current trial and go onto the next one. If the individual guesses correctly, the person is told that he or she is correct, and is presented with a new trial. The score from the PMP is the time spent on impossible items since these times represent effort expended on frustrating tasks.

3.2.2. Game-based assessment of persistence

We created a game-based assessment of persistence (GAP) based on the 74 problems in Newton's Playground. Based on the theory of the PMP (Ventura et al., 2012), the GAP is derived from time spent on unsolved problems over all events in the player's log file over the five sessions. The time spent on each unsolved and solved problem was summed across all events from the log file over the five sessions. For example, if a player attempted (but did not solve) a problem ten different times, the time spent on that problem would be summed across all ten attempts. The average time is then taken for all the unsolved problem sums (out of a possible 74 sums). The same time calculation was made for silver and gold solutions.

3.2.3. Qualitative physics test

Working with a physics professor, we developed a qualitative physics test consisting of 24 pictorial multiple choice items (four choices per item). Its purpose is to assess non-verbal understanding of Newton's three laws, balance, mass, gravity, conservation of momentum, potential and kinetic energy (see Masson et al., 2011; Reiner et al., 2005). We split the qualitative physics test into two forms that were counterbalanced between pretest and posttest (Form A = 12 items; Form B = 12 items).

3.2.4. Self-report measure of persistence

Eight self-report, 5-point Likert-scale items were selected from the International Personality Item Pool, intended to measure perceived persistence on various tasks and in different situations. Care was taken to avoid items that were too broad in nature, so we selected items that targeted performance relative to difficult problems (e.g., *I have patience when it comes to difficult problems, I get easily frustrated on new problems, I tend to avoid difficult problems, I put little time and effort into my work*).

3.2.5. Enjoyment question

A 5-point Likert-scale question at the end of the study was asked about how much students liked NP: *I enjoyed playing Newton's Playground*.

3.2.6. Video game experience

Participants answered one question about general video game use: *How often do you play video games? 1 = not at all, 2 = about once a month, 3 = a few times a month, 4 = a few times a week, 5 = everyday, but for less than 1 h, 6 = every day for 1–3 h, 7 = every day for more than 3 h* (Jackson, Witt, Games, Fitzgerald, von Eye, & Zhao, 2012).

4. Results

Reliability for the physics test was acceptable (Form A: $\alpha = .72$; Form B: $\alpha = .73$). Reliability was good for the PMP ($\alpha = .80$) as well as the self-report persistence items ($\alpha = .83$). Alpha reliability could not be computed for the GAP since players did not solve or even attempt all problems. Additionally, if two players played the same problem one may have solved it while the other may have not solved it, thus the data for the 74 problems was sparse.

Table 1 displays the means for all the measures. Regarding overall learning as a function of NP gameplay, we found a significant difference between the pretest and posttest (t (153) = 2.21, $p < .05$). Over the whole sample the PMP significantly correlates

Table 1
Descriptive statistics for the measures.

	Mean	SD
Pretest	6.23	2.14
Posttest	6.56	2.36
Unsolved	184.75	100.02
Silver	131.75	48.78
Gold	68.15	40.68
Enjoy	3.75	1.11
Self *P*	3.56	0.60

Unsolved = unsolved time; silver = silver time; gold = gold time; enjoy = I enjoyed playing NP; self *P* = self-report measure of persistence.

with the posttest even after controlling for video game experience, gender, physics pretest, and enjoyment ($r = .24$, $p < .01$). The PMP also significantly relates to the unsolved times ($r = .29$, $p < .001$) and silver trophy times ($r = .35$, $p < .001$) but not to gold trophy times. The unsolved times, gold times, and silver times do not relate to NP enjoyment. Finally, the unsolved times, gold times, and silver times do not relate to the posttest after controlling for video game experience, gender, video game use, physics pretest, and NP enjoyment.

To further evaluate the validity of the GAP, we split the sample into low and high performers. NP low performers were identified as those participants who scored in the bottom 50th percentile on number of gold trophies received (i.e., receiving fewer than 9 gold trophies across all sessions). We chose this percentile to ensure we would have an adequate sample. This group of participants did not show significant differences between pretest and posttest mean scores ($p > .05$). Table 2 displays the correlations among the assessments. Based on the correlations between the silver times and the unsolved times we expanded the GAP to include silver times (average of unsolved times and silver times). As can be seen, the GAP, silver, and unsolved times significantly relate to the PMP and the physics posttest scores, but not to NP enjoyment or to the self-reported measure of persistence.

Additionally, the GAP and the PMP relate to the posttest scores of low performers even after controlling for gender, video game experience, physics pretest, and enjoyment ($r = .26$, $p < .05$; $r = .27$, $p < .05$, respectively for GAP and PMP) suggesting that both persistence measures predict learning even after controlling for background knowledge and game enjoyment.

Table 3 displays the correlations among the variables for the 84 high performers in NP. As can be seen, the GAP relates much lower to the PMP for high performers versus low performers. Additionally, the GAP also does not relate to the physics posttest. However the high relation between silver and unsolved times remains high.

Table 2
Correlations for 70 low performers in Newton's Playground.

	PMP	GAP	Unsolved	Silver	Gold	Enjoy	Self-*p*
GAP	.51**						
Unsolved	.47**	.96**					
Silver	.42**	.81**	.62**				
Gold	.00	.22	.14	.32**			
Enjoy	.23	.08	.02	.18	.06		
Self *P*	.10	−.01	−.01	.00	.03	−.05	
Physics post test	.30*	.33**	.31*	.29*	.08	.18	.15

GAP = unsolved and silver times; unsolved = unsolved time; silver = silver time; gold = gold time; enjoy = I enjoyed playing NP; self *P* = self-report measure of persistence.
* $p < .05$.
** $p < .01$.

Table 3
Correlations for 84 high performers in Newton's Playground.

	PMP	GAP	Unsolved	Silver	Gold	Enjoy	Self-*p*
GAP	.22*						
Unsolved	.12	.94**					
Silver	.31**	.84**	.60**				
Gold	.12	.30**	.36**	.11			
Enjoy	.10	.11	.11	.09	−.03		
Self P	.09	−.06	−.05	−.05	−.13	.07	
Physics post test	.16	.02	.01	.04	−.21	.15	.15

GAP = unsolved and silver times; unsolved = unsolved time; silver = silver time; gold = gold time; enjoy = I enjoyed playing NP; self *P* = self-report persistence.
* $p < .05$.
** $p < .01$.

5. Discussion

These results suggest that a valid game-based assessment of persistence can be achieved in a video game. The GAP (unsolved and silver times) was correlated with the PMP, another measure of persistence. The relation between the GAP and PMP measures increased when looking at struggling players in NP. Thus we found evidence of construct validity of the GAP. Both of these measures are grounded on the premise that longer times spent on difficult problems indicate persistence (Eisenberger & Leonard, 1980; Ventura et al., 2012). In addition, the GAP and PMP were positively correlated with the posttest even after controlling for gender, video game experience, pretest knowledge and enjoyment of NP. This suggests that the GAP and PMP measures have criterion related validity (i.e., they both relate to an important outcome – learning qualitative physics). Conversely, the self-reported persistence scores did not predict learning in NP suggesting that it is an inadequate measure for assessing persistence in the context of learning.

The validity of the GAP does appear to depend on whether kids were being sufficiently challenged in the game. That is, players who had more difficulty in the game were operating under the required conditions to elicit persistence (i.e., to persist one must be challenged). This is consistent with the theoretical framework of the PMP which requires students to expend effort on really hard or impossible problems.

One reason why the silver times related to the PMP may be that kids were striving to get a gold trophy and in the process received a silver trophy. Thus the silver times may be conceptually similar to unsolved times in that the silver times represents the overall time spent trying to solve a difficult level (i.e., getting a gold trophy). Finally, the fact that silver times were *positively* correlated to the pretest suggest that the solved times were not a measure of ability but rather persistence (i.e., a negative relation would imply that shorter/faster times were associated with greater knowledge). The lack of relation between gold times and the PMP may be due to the lack of data for gold trophy times considering the low number of gold trophies obtained in the whole sample.

There is a paucity of research in the area of persistence improvement (for a review see Eisenberger, 1992) despite the repeated claims that persistence is a highly valuable skill needed for success in school, on the job, and in life in general (e.g., Roberts et al., 2007). One application of this current work is to use the GAP in real time to build a formative assessment for persistence in NP. This formative assessment could be achieved by assessing persistence and modifying video game difficulty in real time to enhance persistence. For example, if a person is struggling in the game (i.e., not obtaining gold trophies), and not spending adequate time on problems (i.e., the main measure of persistence), NP could alter difficulty and dynamically administer easier problems to try to boost effort. Once the player begins to spend more time on problems, more difficult problems can be presented to simulate greater amounts of persistence. In addition, NP could provide explicit feedback with encouraging prompts (e.g., "Don't give up–you're almost there!"). Implementing such a real-time formative assessment system in NP could be a way to train students to persist despite difficult challenges even outside the context of NP. Future research should also focus on experimental designs comparing video games (e.g., NP) with or without a formative assessment of persistence. Transfer can also be evaluated by seeing if increased persistence in video games affects persistence in other learning contexts.

Additionally, embedding a GAP in NP (or any educational game) could be used to enhance learning gains. In this study we found that the GAP predicts learning gains even after controlling for common predictors of learning (e.g., pretest knowledge, video game use, gender). Thus it is possible to use this persistence information as a measure of engagement to stimulate learning. For example, if a person is not persisting and solving lots of problems then this may be an indicator of boredom in the learner. NP could dynamically introduce more difficult problems to stimulate engagement and challenge the player outside of their range of ability.

The lack of a significant relationship between the self-reported measure of persistence and the GAP could be due to a number of reasons. First, the effects of using different methods of assessment could account for the low correlation between the two measures. A recent study investigating an IAT (implicit association test) of conscientiousness similarly found no relationship between IAT conscientiousness and self-report measures of conscientiousness (Vianello, Robusto, & Anselmi, 2010). Second, performance-based measures may be less susceptible to social desirability effects due to the implicit nature of the PMP assessment (Ventura et al., 2012). Third, the PMP and GAP can be said to have greater face validity than self-report measures of persistence given that the theoretical definition of persistence (i.e., effort expended on difficult tasks) is very similar to the measure itself (time spent on impossible trials/unsolved problems in NP).

6. Conclusion

This study provides evidence of the validity of a game-based assessment of persistence and continues the emerging trend showing that video games and simulations can be useful for assessment purposes (e.g., Quellmalz et al. 2012; Shute & Ventura, 2013). The present data supports the GAP and the PMP as a more valid assessment of persistence than a self-report measure of persistence. Future research should focus on training studies using video games with embedded formative assessment designs to evaluate if persisting in video games can transfer to persistence in other learning contexts.

Acknowledgements

We would like to thank the Bill and Melinda Gates Foundation for their funding of this project. We would also like to thank Matthew Small, Yoon Jeon Kim, and Lubin Wang for their work on this project.

References

Coller, B. D., & Scott, M. J. (2009). Effectiveness of using a video game to teach a course in mechanical engineering. *Computers and Education, 53*(3), 900–912. http://dx.doi.org/10.1016/j.compedu.2009.05.012.

Dede, C. (2005). Planning for neomillennial learning styles. *EDUCAUSE Quarterly, 28*(1), 7–12.

DiCerbo, K. E., & Behrens, J. T. (2012). Implications of the digital ocean on current and future assessment. In R. Lissitz & H. Jiao (Eds.), *Computers and their impact on state assessment: Recent history and predictions for the future* (pp. 273–306). Charlotte, NC: Information Age Publishing.

diSessa, A. A. (1982). Unlearning aristotelian physics: A study of knowledge-based learning. *Cognitive Science, 6*, 37–75.

Dudley, N. M., Orvis, K. A., Lebiecki, J. E., & Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*, 40–57.

Eisenberger, R. (1992). Learned industriousness. *Psychological Review, 99*, 248–267.

Eisenberger, R., & Leonard, J. M. (1980). Effects of conceptual task difficulty on generalized persistence. *American Journal of Psychology, 93*, 285–298.

Hake, R. R. (1998). Interactive engagement vs. traditional methods in mechanics instruction. *American Journal of Physics, 66*(1), 64–74.

Hewitt, P. G. (2009). *Conceptual physics* (11th ed.). San Francisco, CA: Pearson Education.

Jackson, L. A., Witt, E. A., Games, A. I., Fitzgerald, H. E., von Eye, A., & Zhao, Y. (2012). Information technology use and creativity: Findings from the children and technology project. *Computers in Human Behavior, 28*, 370–376.

Lepper, M. R., & Malone, T. W. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. In R. E. Snow & M. J. Farr (Eds.), *Conative and affective process analysis* (pp. 255–286). Hillsdale: Lawrence Erlbaum Associates, Inc..

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.

McCloskey, M., & Kohl, D. (1983). Naive physics: The curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 146–156.

Pausch, R., Gold, R., Skelly, T., & Thiel, D. (1994). What HCI designers can learn from video game designers. In: *Proceedings of the ACM conference on human factors in computing systems* (pp. 177–178). ACM Press.

Perry, S. J., Hunter, E. M., Witt, L. A., & Harris, K. (2010). P = f (Ability × Conscientiousness): Examining the facets of conscientiousness. *Human Performance, 23*, 343–360.

Ploetzner, R., & VanLehn, K. (1997). The acquisition of informal physics knowledge during formal physics training. *Cognition and Instruction, 15*(2), 169–205.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135*, 322–338.

Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). Science assessments for all: Integrating science simulations into balanced state science assessment systems. *Journal of Research in Science Teaching, 49*, 363–393. http://dx.doi.org/10.1002/tea.21005.

Reiner, C., Proffit, D. R., & Salthouse, T. (2005). A psychometric approach to intuitive physics. *Psychonomic Bulletin and Review, 12*, 740–745.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development, 44*(2), 43–58.

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. *Personnel Psychology, 58*, 103–139.

Roberts, B. W., Kuncel, N., Shiner, R. N., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socio-economic status, and cognitive ability for predicting important life outcomes. *Perspectives in Psychological Science, 2*, 313–345.

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

Sweetser, P., & Wyeth, P. (2005). Gameflow: A model for evaluating player enjoyment in games. *ACM Computers in Entertainment, 3*(3), 1–24.

Tobias, S., & Fletcher, J. D. (Eds.). (2011). *Computer games and instruction*. Charlotte, NC: Information Age Publishers.

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

Vianello, M., Robusto, E., & Anselmi, P. (2010). Implicit conscientiousness predicts academic performance. *Personality and Individual Differences, 48*(4), 452–457.

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

White, B. Y. (1994). Designing computer games to help physics students understand Newton's laws of motion. *Cognition and Instruction, 1*(1), 69–108.

Wilson, K. A., Bedwell, W., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J., et al. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation and Gaming, 40*(2), 217–266.