The Journal of Educational Research, 106:423–430, 2013
Copyright © Taylor & Francis Group, LLC
ISSN: 0022-0671 print / 1940-0675 online
DOI:10.1080/00220671.2013.832970

Routledge
Taylor & Francis Group

Assessment and Learning of Qualitative Physics in Newton's Playground

VALERIE J. SHUTE
MATTHEW VENTURA
YOON JEON KIM
Florida State University

ABSTRACT. Digital games are very popular in modern culture. The authors are examining ways to leverage these engaging environments to assess and support student competencies. The authors examine gameplay and learning using a physics game they developed called Newton's Playground. The sample consisted of 167 eighth- and ninth-grade students who played Newton's Playground for about 4 hr over the course of 1.5 weeks. Findings include significant pretest–posttest physics gains, and significant relations between in-game indicators and learning.

Keywords: learning in games, qualitative physics, stealth assessment

In the present article we examine the effectiveness of using a new computer-based game to assess and support learning of physics concepts. Our research in this area has been motivated by a couple of factors. First, schools in the United States have remained virtually unchanged for many decades while our world is changing rapidly (Shute, 2007). As a result, we are seeing a growing number of disengaged students. This disengagement increases the chances of students dropping out of school. For instance, high dropout rates, especially among Hispanic, African American, and Native American students, were described as the silent epidemic in a recent research report for the Bill and Melinda Gates Foundation (Bridgeland, DiIulio, & Morison, 2006). According to this report, nearly one third of all public high school students drop out, and the rate is higher for minority students. In the report, when 467 high school dropouts were asked why they left school, 47% of them simply responded, "The classes were not interesting." We need to find ways, such as using well-designed digital games, to get students engaged, support their learning, and allow them to contribute fully to society. Well-designed games include features that make them intrinsically motivating and thus engaging (Fullerton, 2008; Malone & Lepper, 1987; Shute, Rieber, & Van Eck, 2011). Some of the features of good games include adaptive challenges, goals and rules, interactive problem solving, control (of learning and the game environment), ongoing feedback, and sensory stimuli. We also need to ensure equity in this type of approach, especially given that boys play games more often than girls (e.g., Chou & Tsai, 2007; Lucas & Sherry, 2004).

Another reason for using games as assessments is a pressing need for dynamic and ongoing measures of learning processes and outcomes. Interest in alternative forms of assessment is driven by dissatisfaction with and limitations of multiple-choice items. In the 1990s, interest in alternative forms of assessment increased with the popularization of what became known as authentic assessment. A number of researchers found that multiple-choice and other fixed-response formats substantially narrowed school curricula by emphasizing basic content knowledge and skills within subjects and not assessing higher order thinking skills (e.g., Kellaghan & Madaus, 1991; Shepard, 1991). However, as Madaus and O'Dwyer (1999) argued, incorporating performance assessments into testing programs is difficult because they are less efficient, more difficult and disruptive to administer, and more time consuming than multiple-choice testing programs. Consequently, multiple-choice has remained the dominant format in most K–12 assessments in the United States and elsewhere. New performance assessments are needed that are valid, reliable, and automated in terms of scoring.

In addition to the possibility of using games as assessment devices, there is also growing evidence of video games supporting learning (e.g., Coller & Scott, 2009; Ferguson & Garza, 2011; Tobias & Fletcher, 2011; Wilson et al., 2009). However, learning in games has historically been assessed indirectly and/or in a post hoc manner (Shute & Ke, 2012). We need to understand more precisely how and what kinds of knowledge and skills are being acquired in games. Understanding the relationships between games and learning is complicated by the fact that we do not want to disrupt players' engagement levels during game play. What is needed

---

Address correspondence to Valerie Shute, Instructional Systems Program, Florida State University, 3205G Stone Building, 1114 West Call Street, Tallahassee, FL 32306-4453, USA. (E-mail: vshute@fsu.edu)

instead is real-time assessment and support of learning based on the dynamic needs of players.

Similar to other performance-based assessment in games (see Dede, 2005; DiCerbo & Behrens, 2012; Quellmalz, Timms, Silberglitt, & Buckley, 2012), our stealth assessment approach refers to evidence-based assessments that are woven directly and invisibly into the fabric of the gaming environment (e.g., Shute, 2011). During game play, students naturally produce rich sequences of actions while performing complex tasks. Evidence needed to assess the skills is thus provided by the players' interactions with the game itself. In this article we describe our assessment of qualitative physics understanding in a game we created called Newton's Playground (NP) and examine the effect of gameplay on learning.

Newton's Playground

Research into what is called folk (or qualitative) 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, Proffitt, & Salthouse, 2005). For example, when asked to draw the water level on a picture of a tilted drinking glass, about 40% of young adults draw lines that are not horizontal (McAfee & Proffitt, 1991). When asked to predict the path that a pendulum takes when the string is cut at various points, a large percentage of people make systematically incorrect judgments (Caramazza, McCloskey, & Green, 1981). 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 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 informal 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 direction of a force. Teaching about simple machines (e.g., lever, pulley, pendulum) 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 NP to help secondary school students understand what we call qualitative physics (for details on the design and development of the game, see Shute & Ventura, 2013). We define qualitative physics as a nonverbal conceptual understanding of how the physical world operates, along the lines of Newtonian, not Aristotelian physics. Qualitative physics is characterized by an implicit understanding of Newton's three laws: balance, mass, and conservation and transfer of momentum, gravity, and potential and kinetic energy.

NP is a two-dimensional computer-based game that requires the player to guide a green ball to a red balloon. The player uses the mouse to 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 (which are called agents of force and motion in the game) on the screen with the mouse and colored markers. The objects 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 NP require the player to draw or create four different agents: inclined plane/ramps, pendulums, levers, and springboards. Again, all solutions are drawn with colored markers 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.

For example, in the "golf problem" (see Figure 1), the player must draw a pendulum on a pin (i.e., little circle on the cloud) to make it swing down to hit the ball. In the depicted solution, the player also drew a ramp to prevent the ball from falling down a pit. The speed of (and importantly, the impulse delivered by) the swinging pendulum is dependent on the mass distribution of the club and the angle from which it was dropped to swing. The ball will then fly at a certain speed, length, and trajectory. If drawn properly, the ball will hit the balloon.

Other Gameplay Features

NP consists of seven playgrounds (each one containing around 10–11 levels) that progressively get more difficult. Each level is designed to elicit a particular agent, or in the case of some very difficult levels, a couple of agents. The difficulty of a problem is based on a number of factors including relative location of ball to balloon, number of obstacles present, number of agents required to solve the problem, and novelty of the problem. NP also includes agent tutorial videos that show the player how to create and use the various agents of force and motion. During gameplay, students have the option to watch agent tutorial videos at any time.

Object limit and object use. During pilot testing, we saw that it was possible to game the system by drawing a lot of tiny

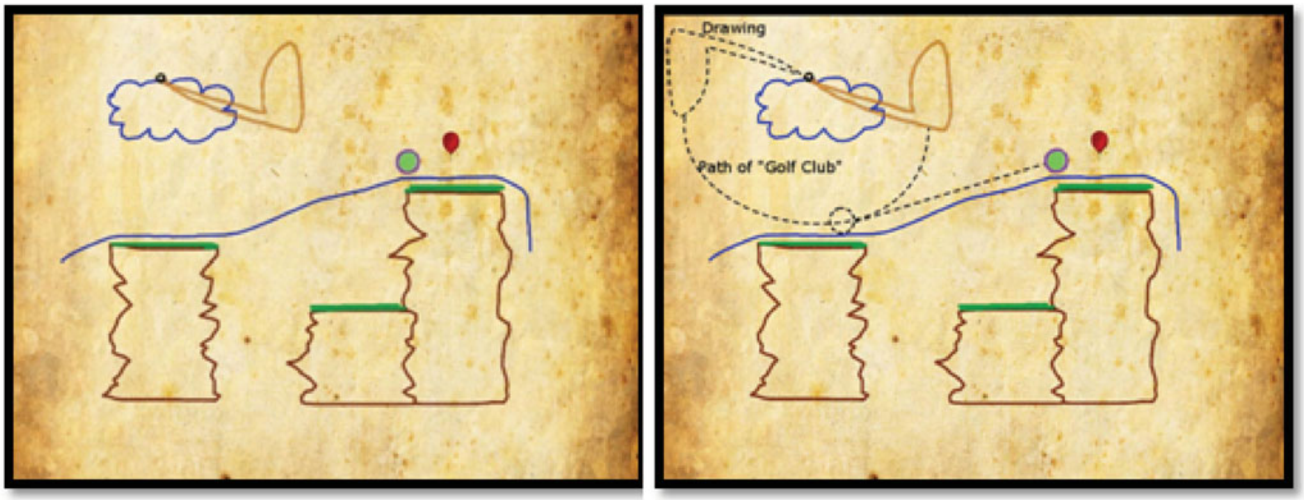


FIGURE 1. Golf problem in Newton's Playground (left is solution; right is path of motion (color figure available online).

objects that stack in some cases to solve a simple problem. We chose to curtail that activity by imposing object limits in the game—where the student had to solve each level using 10 or fewer objects. The log files show on which levels the player hit the object limit, at which point the student has to restart the level or delete objects. This indicator tells us which agent(s) a person is having trouble solving (because we map agents to levels).

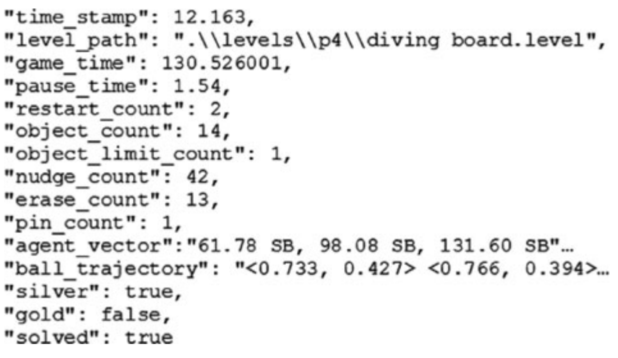
Gold trophies versus silver trophies. If a student solves a level efficiently (i.e., with less than three objects), they receive a gold trophy, and if they solve it with three or more objects, they receive a silver trophy. Gold trophies show that the player has mastered an agent for a given level. Silver means the player may not have fully mastered the agent.

Log Files in NP

The most elemental part of the stealth assessment is the log files generated by NP. NP automatically uploads log files to a server for a session (i.e., log activity between login and logout). Figure 2 shows an example of a player's log file as she's engaged in solving a particular level in the game.

The log file example displays one person's information on one level in the game in terms of particular counts and times for selected features of gameplay (e.g., time spent on the level in seconds, number of restarts of the level, total number of objects used in a solution attempt, whether it was ultimately solved, and trajectory of the ball in the *x*,*y* coordinate space). Each of these variables provides useful information about students' gameplay behaviors, which can then be used to make inferences about how well they are doing in the game and their current understanding of qualitative physics. For example, receipt of gold, silver, or no trophy in a given level suggests the degree of student mastery involving the particular agent of force and motion.

Correct and incorrect use of agents. The log files give detailed information about what correct and incorrect agents were attempted in each level. To accomplish this indicator, we first had to develop an agent identification system in the game, which can make inferences (at about 95% accuracy when compared with human ratings) on what type of agent was drawn by the student. Basically, each agent is characterized by the presence (and absence) of certain features (e.g., pendulum has one pin, an arm, and if the arm moves downward and strikes the ball, it is coded as a pendulum strike by the game). Table 1 displays the rules for agent identification. What students draw in their solution attempts are classified by the gaming system and presented in the log file.



```
"time_stamp": 12.163,
"level_path": ".\\levels\\p4\\diving board.level",
"game_time": 130.526001,
"pause_time": 1.54,
"restart_count": 2,
"object_count": 14,
"object_limit_count": 1,
"nudge_count": 42,
"erase_count": 13,
"pin_count": 1,
"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
```

FIGURE 2. Level log file data.

TABLE 1. Agent Identification System in Newton's Playground

Agent	Monitor Trigger	Identify Trigger
Ramp	Event Ball touches Primary Object Conditions • Object has never rotated >20°	Event Positive ID conditions met OR ball stops touching Primary Object Conditions • Object has never rotated > 20° • Ball moves along object: (>25% in horiz.) OR (>11% horiz. AND >4% vert.) OR (>4% horiz. AND >11% vert.)
Lever	Event Secondary Object falls on Object Conditions • Secondary Object has elevated downward momentum (vertical momentum <−.05 kg m/s) • Object has ≤1 pin (attached to static object) • Object has not moved much recently (less than 2% of screen in 0.33 s)	Event 0.75 s pass from Monitor Trigger Conditions • Object has touched ball since Monitor Trigger • Object has rotated >20° since Monitor Trigger • Ball has reached an apex 4% higher than at Monitor Trigger
Pendulum strike	Event Object touches ball Conditions • Object has 1 pin • Object has rotated >20° • Object has nonzero rotational velocity	Event 0.75 s pass from Monitor Trigger Condition •Ball moved moderately since Monitor Trigger (>15% screen)
Springboard	Event Object has elevated rotational velocity (> 1.5 m/s) Conditions • Rotating toward 12 o'clock (as opposed to 6 o'clock) • Object has 2+ pins (attaches to a static object)	Event 0.75 s pass from Monitor Trigger Conditions • Object has touched ball since Monitor Trigger • Ball has reached an apex 6% higher than at Monitor Trigger

The Present Study

Research on video games and learning typically does not pay attention to specific performance in the game itself (Shute & Ke, 2012). This study aims to show how playing NP can improve students' understanding of qualitative understanding of physics principles in the context of simple machines. Additionally, we examine how performance in NP relates to existing understanding of basic mechanics. Establishing the validity of the stealth assessment in NP lays the foundation for developing diagnostic support mechanisms (e.g., feedback).

We formulated three hypotheses in this study:

Hypothesis 1 (H_1): Overall learning—players would learn qualitative physics as a function of playing NP.

H_2: Engagement—engaged players (as measured by in-game indicators) would show greater learning of qualitative physics compared with less engaged players in NP.

H_3: Validity—performance indicators in NP would relate to existing physics knowledge.

There are two main indicators from the log file that we predict will be related to physics knowledge: (a) number of gold trophies per agent and (b) number of silver trophies per agent.

Additionally, we explored individual differences (i.e., gender, frequency of game play in daily lives, and enjoyment of NP) and examine how those variables affect performance in NP.

Method

Sample

Our sample consisted of 167 eighth- and ninth-grade students (76 male, 91 female). Each student was paid $25 for participation. Students were enrolled at the Florida State University School (FSUS) and tested in groups of about 20 students, per session. Students in Grades 8 and 9 were selected because of the alignment of NP content and the Next Generation Sunshine State Standards (relating to Newtonian Physics) at those grade levels.

Procedure

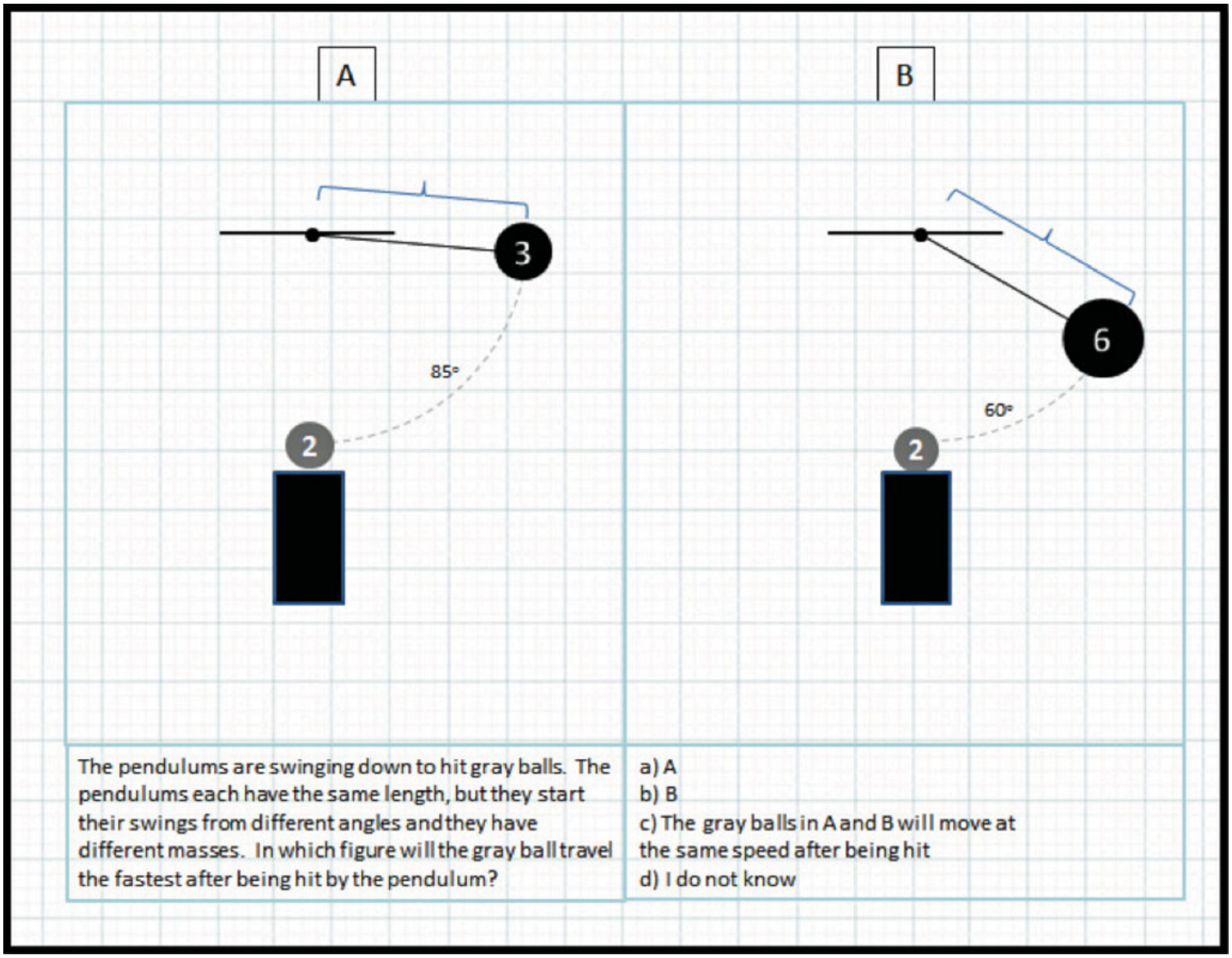
Students played NP for around 4 hr altogether (split into six 45-min sessions that spanned about a week and a half). The study took place in one of the FSUS computer labs. The computer lab contained about 30 computers, each with a monitor, mouse, keyboard, and headphones. Although computers were located next to each other around the lab, each computer was surrounded by carrels to ensure that students do not talk to other students or look at other students' computer screens.

We administered a qualitative physics pretest at the beginning and a posttest at the end of the gameplay sessions (both online). We also administered performance-based measures of creativity and persistence before gameplay, as well as a demographic questionnaire about their age, gender (males = 0, females = 1), frequency of game play, and so on. 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 agent 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 if they were stumped on a problem.

Measures

Qualitative physics test. Working with a physics professor, we developed a qualitative physics test consisting of 24 pictorial multiple-choice items. Its purpose is to assess implicit knowledge of Newton's three laws: balance, mass, and conservation and transfer of momentum, gravity, and 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). For example, Figure 3 shows an item involving a pendulum. The correct answer is B. Reliability for the physics test was acceptable (Form A: Cronbach's $\alpha = .72$; Form B: Cronbach's $\alpha = .73$).

NP enjoyment and frequency of game play. After the students completed NP, we asked them questions about the game. One item asked them to rate the following: "I enjoyed



The pendulums are swinging down to hit gray balls. The pendulums each have the same length, but they start their swings from different angles and they have different masses. In which figure will the gray ball travel the fastest after being hit by the pendulum?

a) A
b) B
c) The gray balls in A and B will move at the same speed after being hit
d) I do not know

FIGURE 3. Example test item from physics test (color figure available online).

TABLE 2. Learning Gains and Engagement in Newton's Playground

Levels attempted in Newton's Playground/ Engagement	Pretest		Posttest	
	M	*SD*	*M*	*SD*
Low ($n = 47$)	5.77	2.39	5.81	2.55
Medium ($n = 52$)	6.52	2.17	7.00	2.11
High ($n = 55$)*	6.35	1.98	6.87	2.22

*$p < .05$.

playing Newton's Playground," rated on a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). At the beginning of the study, we asked students about the frequency of game play in their daily lives: "How often do you play video games?" This was rated on a 7-point Likert-type scale ranging from 1 (*I never play*) to 7 (*I play > 3 hr a day*).

Results

Overall Learning

Regarding learning physics as a function of simply playing NP, we found a significant difference between the pretest and posttest scores, $t(154) = 2.12$, $p < .05$. Students playing the game improved in their qualitative, conceptual physics understanding over time.

Engagement and learning. To understand how engagement influenced learning in NP, we analyzed learning gain by number of levels attempted in NP as an indicator for engagement (high, medium, and low based on cumulative percentiles ranging from 0 to 74). The results of the analysis of variance (ANOVA) showed no significant pretest differences relative to level of engagement, $F(2, 163) = 1.64$, $p = ns$, but significant posttest differences by level, $F(2, 151) = 3.98$; $p = .02$. Table 2 displays the means (and standard deviations) of the pretest and posttest as a function of levels attempted. The students who elected to play more levels (i.e., the high-engagement group) showed significant gains from pretest to posttest, $t(54) = 2.36$, $p = .02$. The medium-engagement group showed marginal gains, $t(51) = 1.86$, $p = .07$, and there was no significant difference between pretest and posttest for the low-engagement students, $t(46) = 0.11$, $p = ns$.

Validity

We hypothesized that certain performance indicators in NP would relate to existing physics knowledge—specifically the number of silver and gold trophies obtained throughout gameplay. Significant relations would suggest construct validity. Table 3 displays the correlations among the number of silver and gold trophies received during gameplay and pretest knowledge. Note that all of the gold trophies per agent relate significantly to pretest data, but only a couple of the silver trophies do. This is in line with our belief that getting gold trophies shows that the player has mastered an agent for a given level while silver means the player may not have fully mastered the agent.

Individual Differences

Finally, we explored how several individual differences measures related to performance in NP. Overall, the sample enjoyed playing NP ($M = 3.8$, $SD = 1.1$; see Figure 4). However, NP enjoyment does relate to gender ($r = -.17$, $p < .05$), suggesting that males tend to like NP more than females. Males also showed higher pretest scores than females, which might have affected NP enjoyment. To address this issue, we examined gender-related variables more closely while controlling for pretest knowledge.

Table 4 displays the partial correlations (controlling for pretest) among gender, personal video game use, gold trophies obtained in NP, and enjoyment of NP. As can be seen, gender does not relate to enjoyment after controlling for pretest knowledge. Moreover, a person's gameplay experience does not affect enjoyment.

Do males in our sample learn more qualitative physics from NP compared with females? We computed a repeated-measures ANOVA on pretest and posttest data, using gender as our between-subjects factor. As noted earlier, there was a main effect of overall learning, $F(1, 153)$

TABLE 3. Correlations Between Pretest Scores and Newton's Playground Trophies (N = 166)

	Posttest	RAs	LEs	PEs	SBs	RAg	LEg	PEg	SBg
Pretest	0.60**	0.16*	0.03	0.06	0.29**	0.27**	0.22**	0.31**	0.40**

Note. RA = ramp; LE = lever; PE = pendulum; SB = springboard; s = silver; g = gold.
*$p < .05$. **$p < .01$.

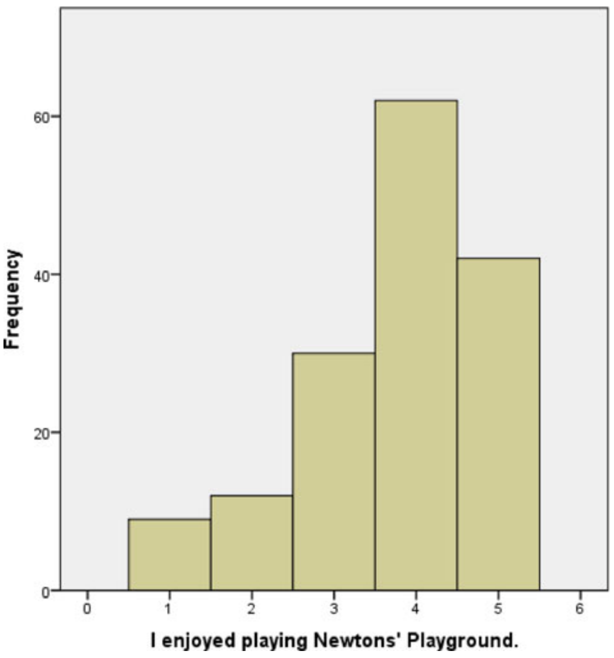


FIGURE 4. Enjoyment of Newton's Playground ratings on a 5-point Likert-type scale (color figure available online).

= 4.24, $p < .05$, however there was no significant gender by learning interaction, $F(1, 153) = 0.63$, $p = ns$. Table 5 shows that while males began the experiment with greater physics knowledge than females, both males and females showed comparable and significant learning over time.

Discussion

In this study we aimed to explore the effectiveness of NP as an assessment system for qualitative, conceptual physics understanding as well as its potential influence on learning physics in a game without any instructional support. First, we showed that playing NP for about 4 hr across 1.5 weeks can lead to improved understanding of qualitative physics knowledge, so our first hypothesis was supported. Playing a video game that requires use of simple machines can indeed reinforce qualitative physics understanding as measured by an external assessment.

Next, we wanted to test the effects of differential engagement in the game (as assessed by number of levels attempted) on learning. Our second hypothesis was also confirmed in that students with greater levels of engagement showed significant pretest to posttest gains, while those with low levels of engagement did not demonstrate learning gains.

We also found that particular actions and accomplishments within the game can be used for assessment purposes. Specifically, in-game performance related to creating and using various agents in NP correlated to existing physics knowledge (i.e., as measured in the pretest). That is, attaining gold trophies for a given agent appears to be a better indicator of physics knowledge than silver trophies. This is understandable because attainment of a gold trophy usually requires the proper use of agents using just a few objects, and thus our third hypothesis was also supported.

TABLE 4. Correlations Controlling for Pretest Knowledge

	Game	Gold	Enjoy
Gender	−0.38**	−0.47**	−0.12
Game		0.27**	0.14
Gold			0.21*

Note. Game = Game play frequency; Gold = gold trophies obtained; Enjoy = enjoyment in Newton's Playground.
$^*p < .05$. $^{**}p < .01$.

Regarding individual differences related to gameplay and learning, we found that while boys arrived at the experiment with more incoming physics knowledge than females in our sample, both improved comparably from pretest to posttest. This shows that NP is not biased in favor of particular populations. We also found that students generally enjoyed NP (albeit, males slightly more than females). However, after controlling for pretest knowledge, NP enjoyment was not related to either gender or to history of video game use.

Looking forward, stealth assessment as used in NP has the potential to be useful for diagnostic and support purposes. For example, if a student has trouble using a particular agent, certain gameplay features could inform the most likely reasons why that's the case. For instance, a player's lever solution may have failed because (a) the wrong mass of an object was used on one side of the lever, (b) the fulcrum was positioned inaccurately, and/or (c) the size or length of the lever was too short or too long. Based on this information, NP can give feedback as to how to correctly draw agents of force and motion.

Additional research in the area of stealth assessment includes working with teachers to embed NP into the physical science curriculum. This would involve linking Newtonian

TABLE 5. Pretest and Posttest Means and Standard Deviations by Gender

		M	*SD*
Pretest	Male ($n = 76$)	6.67	2.01
	Female ($n = 91$)	5.86	2.17
Posttest	Male ($n = 72$)	6.90	2.43
	Female ($n = 83$)	6.27	2.27

physics formalizations (e.g., F = ma) to relevant NP problems for instructional support. Teachers can also design their own levels in NP to highlight physics concepts that could benefit from more hands-on experience and support. Additional scaling of the game includes (a) adding more levels to game, especially interactions among Newton's laws of motion; (b) creating more physics content, like principles of collision; (c) examining predictive validity of the game relative to future science courses taken and grades received therein; (d) using the indicators associated with the four agents of force and motion to infer misconceptions for diagnostic and support purposes; and (e) expanding the platform of NP from computer- to browser-based gameplay.

We are excited that researchers are starting to use digital games for learning and assessment. We think stealth assessment is one way to maximize the positive impact digital games can have on students.

ACKNOWLEDGMENTS

The authors would like to thank the Bill and Melinda Gates Foundation for their funding of this project. They would also like to thank Matthew Small, Lubin Wang, and Don Franceschetti for their work on this project.

REFERENCES

Bridgeland, J. M., DiIulio, J. J., Jr., & Morison, K. B. (2006). *The silent epidemic: Perspectives of high school dropouts*. Washington, DC: Civic Enterprises in association with Peter D. Hart Research Associates for the Bill and Melinda Gates Foundation.

Caramazza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in "sophisticated" subjects: Misconceptions about trajectories of objects. *Cognition*, *9*, 117–123.

Chou, C., & Tsai, M.-J. (2007). Gender differences in Taiwan high school students' computer game playing. *Computers in Human Behavior*, *23*, 812–824.

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

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

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

Ferguson, C. J., & Garza, A. (2011). Call of (civic) duty: Action games and civic behavior in a large sample of youth. *Computers in Human Behavior*, *27*, 770–775.

Fullerton, T. (2008) *Game design workshop: A playcentric approach to creating innovative games*. Burlington, MA: Elsevier.

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

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

Kellaghan, T., & Madaus, G. F. (1991). National testing: Lessons for America from Europe. *Educational Leadership*, *49*, 87–93.

Lucas, K., & Sherry, J. L. (2004). Sex differences in video game play: a communication-based explanation. *Communication Research*, *31*, 499–523.

Madaus, G. F., & O'Dwyer, L. M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappan*, *80*, 688–695.

Malone, T. W., & Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning, and instruction* (Vol. 3, pp. 223–253): Hillsdale, NJ: Erlbaum.

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.

McAfee, E. A., & Proffitt, D. R. (1991). Understanding the surface orientation of liquids. *Cognitive Psychology*, *23*, 483–514.

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, & Cognition*, *9*, 146–156.

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.

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

Shepard, L. (1991). Will national tests improve student learning? *Phi Delta Kappan*, *73*, 232–238.

Shute, V. J. (2007). Tensions, trends, tools, and technologies: Time for an educational sea change. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 139–187). New York, NY: Erlbaum.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, NC: Information Age.

Shute, V. J., & Ke, F. (2012). Games, learning, and assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 43–58). New York, NY: Springer.

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.

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

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

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

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

AUTHORS NOTE

Valerie J. Shute is the Mack & Effie Campbell Tyner Endowed Professor in Education in the Department of Educational Psychology and Learning Systems at Florida State University. Her current research involves using immersive games with stealth assessment to support learning—of cognitive and noncognitive knowledge, skills, and dispositions. Her research has resulted in numerous grants, journal articles, books, chapters in edited books, and a patent.

Matthew Ventura is a Senior Research Scientist in the College of Education at Florida State University. His research focuses on the design of educational video games. His work has resulted in numerous publications in the areas of assessment and educational psychology.

Yoon Jeon Kim is a doctoral candidate in the Department of Educational Psychology and Learning Systems at Florida State University working with her advisor, Dr. Valerie Shute, developing stealth assessments in video game environments. Her research interests focus on assessment of complex skills and data mining.