Chapter 8 – Assessing Persistence in Educational Games

Matthew Ventura[1], Valerie Shute[1], and Matthew Small[2]
[1]Florida State University; [2]Independent researcher

Introduction

In this chapter, we describe how assessing persistence in educational games can be useful in enhancing learning. We describe Newton's Playground (NP), a 2D video game aimed at helping students understand qualitative physics. The goal across all of the 74 puzzles in NP is for the player to guide a green ball to a red balloon. The primary way to move the ball is by drawing on the screen with a mouse. Once objects are drawn they are affected by gravity and Newton's three laws of motion. We describe results from a study where we mined data from NP log files to develop an in-game measure of persistence and report on how persistence in NP predicts learning gains. Finally, we propose various ways to use the detection of persistence to enhance learning in educational games using the GIFT framework.

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). Playing digital games has been shown to be positively related to a variety of cognitive skills (e.g., visual-spatial abilities, Green & Bavelier, 2007; Ventura, Shute, Wright & Zhao, 2013; attention, Shaw, Grayson & Lewis, 2005), personality traits (e.g., Openness, Chory & Goodboy, 2011; Ventura, Shute & Kim, 2012; Witt, Massman & Jackson, 2011), persistence (Ventura, Shute & Zhao, 2012), academic performance (e.g., Skoric, Teo & Neo, 2009; Ventura, Shute, Kim, 2012), and civic engagement (Ferguson & Garza, 2011). Moreover, educational games have been shown to enhance learning of academic content, within and outside of the game (e.g., Barab, Dodge et al., 2010; Coller & Scott, 2009; DeRouin-Jessen, 2008).

In addition to video games' effects on learning, they produce a vast amount of data that can be used for assessment purposes (Dede, 2005; DiCerbo & Behrens, 2012; Quellmalz, Timms, Silberglitt & Buckley 2012). Using this stream of data, formative assessments that are embedded in games can enable us to more accurately provide feedback and change gameplay to maximize learning according to the ability level of the player. One way to meet these requirements is to use *stealth assessment* (Shute, 2011). Stealth assessment refers to assessments that are woven directly and invisibly into the fabric of the gaming environment. During game play, students naturally produce rich sequences of actions while performing complex tasks, drawing on the very skills or competencies that we want to assess (e.g., scientific inquiry skills, creativity). Evidence needed to assess the skills is thus provided by the players' interactions with the game itself, which can be contrasted with a typically singular outcome of an activity – the norm in educational environments.

Making use of this stream of gameplay evidence to assess students' knowledge, skills, and understanding (as well as beliefs, feelings, and other learner states and traits) presents problems for traditional measurement models used in assessment. First, in traditional tests the answer to each question is seen as an independent data point. In contrast, the individual actions within a sequence of interactions in a game are often highly dependent on one another. For example, what one does in a particular game at one point in time affects subsequent actions later on. By analyzing a sequence of actions within a quest (where each response or action provides incremental evidence about the current mastery of a specific fact, concept, or skill), stealth assessments within game environments can infer what learners know and do not know at any point in time.

The main assumptions underlying educational games include the following: (a) learning by doing (required in game play) improves learning processes and outcomes, (b) different types of learning and learner attributes may be verified and measured during game play, (c) strengths and weaknesses of the learner may be capitalized on and bolstered, respectively, to improve learning, and (d) ongoing feedback can be used to further support student learning. In line with these assumptions, a central question in this chapter is, *Can we enhance learning in educational games through the detection of persistence?* 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; Ventura, Shute & Zhao, 2012). 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). 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 design principle of challenge in well-designed games (Pausch, Gold, Skelly & Thiel, 1994). That is, in games, establishing the right level of challenge entails adjusting the optimal level of difficulty for a player. This balance – between game difficulty and player ability level – is consistent with the theory of the ZPD (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).

To illustrate this idea of repeated exposure to challenge and the relationship to persistence, Eisenberger and Leonard (1980) showed that trying to complete difficult tasks can improve persistence. Participants were randomly assigned to solve impossible, hard, or easy anagrams and then take the perceptual comparison task where they were asked to detect as many differences as possible between two pictures. Participants who had experienced 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 effort on subsequent tasks. The next section introduces a video game we developed that requires persistence due to its difficulty.

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 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 computer video game called Newton's Playground (Shute & Ventura, 2013) to help middle school students better understand *qualitative physics* (Ploetzner & VanLehn, 1997). Qualitative physics is a nonverbal understanding of Newton's three laws, balance, mass, gravity, conservation of momentum, potential, and kinetic energy. NP is a nonlinear video game (i.e., a game design feature that allows many possible solution paths) where players are challenged to guide a green ball to a red balloon in a 2D environment, where all objects obey the basic rules of physics relating to gravity and Newton's three laws of motion. The player can add a slight impulse to the ball by clicking on it, but the primary way to guide the ball is by drawing simple machines on the screen that "come to life" once the mouse button is released (a gameplay mechanic inspired from games such as Magic Pen and Crayon Physics Deluxe). Using this drawing mechanic, players can create any shape imaginable to help move the ball toward the goal and solve the level. Players can also use pins, which act as a rotating joint, to attach objects to one another.

The 74 problems in NP require the application of four categories of simple machines: inclined plane/ramps, levers, pendulums, and springboards. These simple machines, which we often refer to as "agents of force and motion," or just "agents," are created by a combination the player drawing colored lines on the screen in various shapes (i.e., objects) and attaching objects to each other using pins. A ramp is an object that helps to guide a ball in motion. It can be useful to transform vertical motion to horizontal motion (and vice versa) or guide the ball over a hole. A lever rotates around a fixed fulcrum or pivot point and is generally useful when a player wants to move the ball vertically. The rotation of the lever can be generated by the weight distribution of the lever itself or by an external object transferring momentum to the lever on one side of the fulcrum. A pendulum swings on a pin and directs an impulse tangent to its direction of motion. With enough space, a pendulum can be used to exert a horizontal or vertical force. A springboard (or diving board) stores elastic potential energy provided by a falling weight. Springboards provide an efficient mechanism to move the ball vertically.

Figure 1 displays a puzzle in NP. In this puzzle, 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.

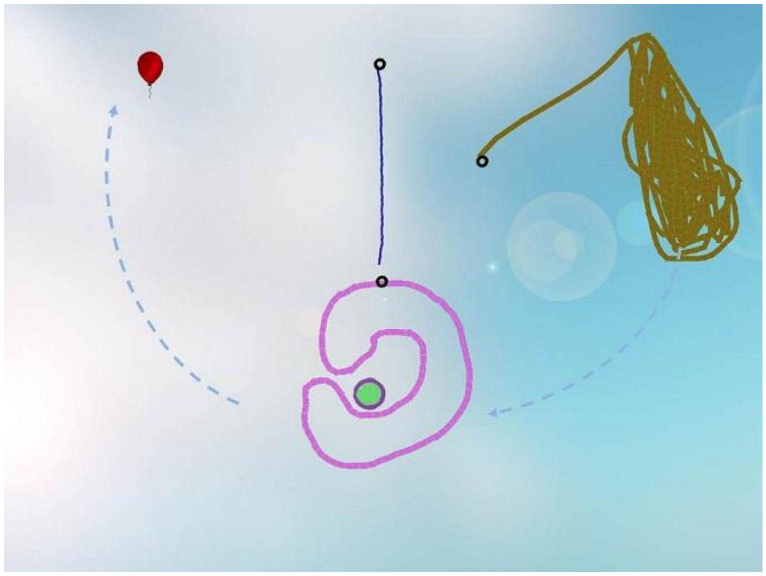

Figure 1. Example problem in NP.

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 set of simple machines (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.

NP Session Logs

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

```
"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
```


Figure 2. NP session event log.

The Impact of Persistence on Learning Gains in Newton's Playground

Ventura and Shute (2013) analyzed log files from 70 8th and 9th grade students who played NP for around four hours (split into five 45-minute sessions across two weeks) in a large computer lab. Based on the theory of the persistence (Ventura, Shute, Zhao, 2012), we developed a game-based assessment of persistence (GAP), which is derived from time spent on *unsolved problems* over all events in the player's log file over the five sessions. That is, longer times spent on difficult problems (whether they were solved or not) should indicate greater persistence (Eisenberger & Leonard, 1980; Ventura, Shute & Zhao, 2012). Time on solved problems should not be an indicator of persistence since solution times are primarily based on skill in the game. The time spent on each unsolved 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 10 different times, the time spent on that problem would be summed across all 10 attempts. The average time is then taken for all the unsolved problem sums (out of a possible 74 problems).

As part of our validation process, we also administered another performance based measure of persistence consisting of impossible anagrams and picture comparison tasks (see Ventura & Shute, 2013). Impossible anagrams consist of jumbled letters that do not actually make a word. Impossible picture comparison items consist of two adjacent pictures where participants are told to detect difference between pictures when in fact no differences exist. 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, and is presented with a new trial. The score from the anagrams and picture comparison tests is the time spent on impossible trials since these times represent effort expended on frustrating tasks.

Ventura and Shute (2013) found that the GAP (i.e., unsolved times) predicted a number of learning gains for struggling players in NP. Table 1 displays the correlations among the GAP and other measures of persistence and learning. As can be seen, the GAP significantly relates to the anagrams and picture comparison tasks and the physics post-test scores. Additionally, the GAP relates to the post-test scores of low performers even after controlling for gender, video game experience, physics pretest, and enjoyment ($pr = 0.26$, $p < 0.05$) suggesting that both persistence measures predict learning even after controlling for background knowledge and game enjoyment.

Table 1. Correlations among performance measures in NP (Ventura & Shute, 2013).

	A-PC	GAP	Gold
GAP	0.47**		
Gold	0.00	0.14	
Physics post-test	0.30*	0.31*	0.08

* = p < 0.05; ** = p < 0.01; A-PC = anagrams and picture comparison task; GAP = unsolved times

The results from Ventura and Shute (2013) suggest that persistence can positively impact learning for struggling students. The GAP was positively correlated with the posttest even after controlling for gender, video game experience, pretest knowledge, and enjoyment of NP. This suggests that the GAP plays an important role in learning. Additionally, the GAP was correlated with the anagrams and picture comparison tasks, another measure of persistence. Thus we found evidence of construct validity of the GAP. Both of these persistence measures are grounded on the premise that longer times spent on difficult problems indicate persistence (Eisenberger & Leonard, 1980; Ventura, Shute & Zhao, 2012).

The usefulness 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 persistence which requires students to expend effort on really hard or impossible problems.

Enhance Learning in Games with GAP

Despite 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, Kuncel, Shiner, Caspi & Goldberg, 2007), there is no prior empirical research testing the relationship between persistence and learning in educational games. Ventura and Shute (2013), however, provide preliminary evidence that a relation exists between persistence and learning. We believe that future work should explore this relationship further, as well as other ways to enhance learning in games by integrating GAP into educational game design. Next, we discuss three potential methods to apply GAP data in educational game design: tuning gameplay difficulty, hints, and feedback.

A primary goal of using GAP to inform the tuning of gameplay difficulty is to keep players in the ZPD, where they are challenged enough to exhibit persistence (and experience its potential benefits) but are not so challenged that they are unable to complete any game objectives or experience excessive frustration. The tuning itself can be performed nearly continuously during gameplay (e.g., adjustments to the quantity and strength of generated enemies) or at a larger quantum (e.g., serving more difficult levels). Optimally selecting the granularity of tuning is largely dependent on game structure and the granularity of GAP data points. We plan to develop additional fine grained data points for NP that inform GAP calculations more continuously (e.g., in-game drawing data), but these techniques will likely be more specific to the game mechanics of NP and thus less transferable to other games.

Hints are another potential target for exploiting GAP within game design to enhance learning. Hints can provide an effective resource for guiding players through difficult, unfamiliar, or poorly understood game scenarios. A constraint to the overall efficacy of hint generating systems (e.g., ITSs) is access to informative data points about the current state of the player’s skills and conceptualizations. Since a primary goal of a hint generating system is to provide players with minimal guidance to complete game objectives (i.e., keep players in the ZPD), GAP could prove an effective tool in producing more targeted hints.

We are also interested in the potential benefits of providing players with explicit and real-time feedback about their GAP level. By providing players with such information regarding their persistence, we are essentially providing the player with a new metric by which to measure their gameplay performance, like a badge or trophy. This metric could become a tool for players to self-regulate their persistence and potentially learn to become more persistent by doing so. Designing GAP feedback would need to be done with care to both give players some indication of potential strategies for regulating their GAP and avoid exposing so much about GAP calculations as to allow “gaming of the system.”

Persistence in the Generalized Intelligent Framework for Tutoring

We have outlined a method to detect persistence demonstrated within educational games. One advantage of the simplicity of the data needed to detect persistence is that it can be applied to a variety of educational products. The main aim of GIFT is to support varying open and dynamic game-based learning environments that apply distinctively different messaging protocols. This involves embedding components and processes within GIFT’s domain module to support the detection of persistence regardless of the educational game being utilized.

GIFT requires rules and models built around game interaction that must be explicitly linked to concepts defined inside of the GIFT architecture. For this purpose, a ‘Gateway Module’ is incorporated that associates an external educational/training system’s state data with a domain or competency model built within GIFT. This linkage allows for two disparate systems to communicate with one another. In the case of persistence within GIFT, this enables the application of real-time assessment of persistence in players.

This allows any system to link interaction with GIFT’s domain model, where persistence assessments are conducted and progress is communicated to the learner model for determining transitions in performance or competency. This approach to assessment is ideal in game-based environments as tracking interaction data as it relates to objectives can denote comprehension and understanding that is difficult to gauge in traditional assessment techniques. The application of stealth assessment within GIFT potentially provides further diagnosis of game performance, which can be communicated to the pedagogical model for more focused selection of feedback and remediation tactics.

References

Coller, B. D. & Scott, M. J. (2009). Effectiveness of using a video game to teach a course in mechanical engineering. *Computers & Education, 53*(3), 900-912.

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 futur*e (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.

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.

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.

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 & Cognition*, 9, 146–156.

Pausch, R., Gold, R., Skelly, T. & Thiel, D. (1994). What HCI designers can learn from video game designers. *Proceedings of the ACM Conference on Human Factors in Computing Systems* (pp. 177-178). ACM Press.

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

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

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. and 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. doi: 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., 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.

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.

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 Publishers.

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. J. (2013). The validity of a game-based assessment of persistence. Computers in Human Behavior, 29, 2568-2572.

Ventura, M., Shute, V., Wright, T. & Zhao, W. (2013). An investigation of the validity of the virtual spatial navigation assessment. *Frontiers in Psychology, 4:852.*

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.

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., … Conkey, C. (2009). Relationships between game attributes and learning outcomes: Review and research proposals. *Simulation & Gaming, 40*(2), 217–266.