How to Increase Learning while not Decreasing the Fun in Educational Games

Valerie Shute, Fengfeng Ke, Russell Almond, Seyedahmad Rahimi, Ginny Smith, & Xi Lu
Florida State University

Abstract

In this chapter we discuss a next-generation learning game—Physics Playground—that successfully blurs the distinction between assessment and learning to promote STEM competency development. We are using evidence-centered design to integrate game-based learning, problem-based learning, personalized learning, and learning by design; linking the activities to both informal and formal physics knowledge. Our focus in this chapter is on the design of various in-game learning supports that offer just-in-time explanations when students succeed, and encouragement and instructional scaffolding when they struggle. Our current learning supports include worked examples, physics animations, constructed definitions, short videos, and relevant formulas. This chapter informs researchers in computer and learning sciences on the design and development of effective in-game learning supports, and the methodology of data mining that can potentially drive a dynamic delivery of learning supports. Moreover, this chapter will inform science educators in developing effective education programs for youth.

Introduction

“Fun from games arises out of mastery. It arises out of comprehension. It is the act of solving puzzles that makes games fun. In other words, with games, learning is the drug.”—Raph Koster

The big problem that we are tackling with the research discussed in this chapter is that the U.S. has been a global leader in STEM-related areas, but this top-tier position is currently threatened by a decreasing number of U.S. students choosing to pursue expertise in STEM fields and an inadequate supply of teachers skilled in those subjects (U.S. Department of Education [USDOE], 2015). According to the President’s Council of Advisors on Science and Technology (2012), economic projections indicate the need for about 1,000,000 more STEM professionals than we will actually produce over the next decade. Unfortunately, the math and science data of our students suggests we are not on track to reach our projected needs. The U.S. now lags behind other nations in STEM education at the elementary and secondary levels (Hanushek, Peterson, & Woessman, 2012; PISA 2012; TIMSS 2011).

To keep the U.S. competitive in the global economy, we need to increase the number and diversity of students entering into STEM areas. Our work, described in this chapter, focuses on achieving two goals that could have the largest impact on meeting this challenge. The first goal is to get more children excited about and interested in science—specifically physics. Recognizing that interest alone is not enough, our second goal is to identify ways to facilitate and deepen science-related learning in immersive and meaningful learning environments. Well-designed digital games
represent a promising vehicle for meeting both goals: capturing children’s interest in STEM fields like physics, and supporting active, contextualized learning.

Review of Relevant Literature

Digital Games in Education

Besides being a popular activity across all gender, ethnic, and socioeconomic lines, playing digital games has been shown to be positively related to various competencies, attributes, and outcomes such as visual-spatial abilities and attention (e.g., Green & Bavelier, 2007, 2012; Shute, Ventura, & Ke, 2015), openness to experience (Chory & Goodboy, 2011; Ventura, Shute, & Kim, 2012; Witt, Massman, & Jackson, 2011), college grades (Skoric, Teo, & Neo, 2009; Ventura et al., 2012), persistence (Ventura, Shute, & Zhao, 2012), creativity (Jackson et al., 2012), and civic engagement (Ferguson & Garza, 2011). Digital games can also motivate students to learn valuable academic content and skills (e.g., Coller & Scott, 2009; DeRouin-Jessen, 2008; for recent reviews, see Clark, Tanner-Smith, & Killingsworth, 2015; Tobias & Fletcher, 2011; Wilson et al., 2009; Young et al., 2012). However, peer-reviewed literature published on the design and evaluation of game-based learning for science is still limited (National Research Council [NRC], 2011; Young et al., 2012).

We are working to expand the research on game-based learning of science through our ongoing enhancements to a physics game we developed called Physics Playground (PP) (Shute & Ventura, 2013). PP is a computer game that dynamically assesses and supports students’ understanding of qualitative physics (the nonverbal understanding of Newton’s three laws of motion, potential and kinetic energy, balance, mass, and gravity; see Ploetzner & VanLehn, 1997). We believe that PP can leverage the popularity of digital games to capture and sustain students’ attention and teach physics to a broader audience than traditional physics classrooms, making it an attractive tool for teaching physics to diverse gender, ethnic, or economic groups. Next, is the theoretical foundation of learning and motivation, on which we base our research.

Motivation and Learning Support Via Games

There is a convergence between the core elements of a good game and the characteristics of productive learning (Shute, Rieber, & Van Eck, 2011). Our proposition is that (a) learning is at its best when it is active, goal-oriented, contextualized, and interesting (e.g., Bransford, Brown, & Cocking, 2000); and (b) learning environments should be interactive, provide ongoing feedback, grab and hold attention, and have appropriate and adaptive levels of challenge—all features of good games. Gee (2003) has argued that the secret of a good game is not its 3D graphics and other bells and whistles, but its underlying architecture in which each level dances around the outer limits of the player’s abilities, (see also Csikszentmihalyi, 1990, on flow theory). Along the same line, psychologists (e.g., Vygotsky, 1987) have long argued that the best instruction hovers at the boundary of a student’s competence. Finally, both well-designed games and productive learning processes employ ongoing feedback as a major mechanism of play/learning support.

**Motivation support via games.** Well-designed games are highly engaging (e.g., Desurvire, Caplan, & Toth, 2004; Fullerton, Swain, & Hoffman, 2008). Play is voluntary, intrinsically motivating, and involves active cognitive and/or physical engagement that allows for the freedom to fail (and recover) and to experiment freely (Klopfer, Osterweil, & Salen, 2009; Pellegrini, 1995; Rieber, 1996). Unlike “free play,” a game is usually a contest of physical or mental skills and
strengths, requiring the player to follow a specific set of rules to attain a goal (Hogle, 1996). The logic model of the PP design uses core game actions and aesthetics (e.g., puzzle solving, drawing-based creative play, adaptive challenges, and dynamic feedback) to engender motivation, which, in turn, will support game-task engagement and ultimately learning. Specifically, puzzle solving and drawing-based creative play are particularly appealing to female gamers and are gender inclusive (Kinzie & Joseph, 2008; Steiner, Kickmeier-Rust, & Albert, 2009). Adaptive challenges and dynamic performance feedback in a game help to create a sense of competence and an optimal environment for diverse players, which will foster the sense of flow and potentially cultivate the growth mindset that engenders effort-driven, challenge-centered competency development (Dweck, 2006).

**Learning support via games.** Well-designed games can be seen as vehicles for engaging players in iterative intellectual activities. People who want to excel at something—from surgeons to artists—spend countless hours of intellectual effort while practicing their craft. There is considerable evidence in the literature, going back more than 100 years, supporting that practice substantially improves knowledge and skills (e.g., Ericsson, Krampe, & Tesch-Römer, 1993; Newell & Rosenbloom, 1981; Schneider & Shiffrin, 1977; Thorndike, 1898). But practice can be tedious and frustrating, causing some learners to abandon their practice and, hence, expertise development. However, good games can provide an engaging and authentic environment designed to keep practice meaningful and personally relevant. With simulated visualization, authentic problem solving, and instant feedback, computer games can afford a realistic framework for experimentation and situated understanding, and thus act as rich primers for active learning (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005; Gee, 2003; NRC, 2011; Squire, 2006). Furthermore, within-game learning support enables learners to do more advanced intellectual activities and to engage in more advanced thinking than they could without such help (Vygotsky, 1987). The complicated part about including learning support in games is to not disrupt engagement while reinforcing the construction and application of cognitive generalizations that deepen learning and engender transfer to other contexts.

**Types of Learning Supports in Games**

Educational game researchers (e.g., Wouters & van Oostendorp, 2013) have concluded that to keep novice players engaged with and learning from the game, it must include purposively designed learning supports. That is, digital games are complex and challenging environments that demand a lot of cognitive effort, and learners will likely get frustrated, disengaged, or distracted by play without being involved in learning (Wouters, van Nimwegen, van Oostendorp, & van Der Spek, 2013). In that case, learning outcomes may be in jeopardy. Including supports in educational games increases the odds of improving game-based learning engagement and knowledge development.

Wouters and van Oostendorp (2013) conducted a meta-analysis on the effectiveness of various learning supports in educational games. They selected 29 studies (with 3,675 participants) and computed 107 pairwise comparisons to investigate the effectiveness of learning supports in educational games. They found a positive and moderately-weighted effect size of $d = .34$ ($z = 7.26$, $p < .001$) which suggests that the use of learning supports in games can, in fact, improve learning. Furthermore, Wouters and van Oostendorp identified 24 different types of learning supports.

According to Wouters and van Oostendorp (2013), there are eight different types of supports that are most commonly used in educational games: reflection, modeling, advice, collaboration, interactivity, narrative elements, feedback, and modality. The first type of support is
reflection which aims to stimulate learners’ thinking about their performance and learning in the game. Research has shown that knowledge retention is improved if students are required to reflect on what they learned (e.g., Leemkuil, 2006). Some of the learning supports in games categorized under reflection include: (1) self-explanation (asking learners to explain to themselves—verbally or written—as they study a lesson/concept; Johnson & Mayer, 2010), (2) elaboration (extra task-related cognitive activities; Shebilske, Goettl, Corrington, & Day, 1999), and (3) reflective inquiries (e.g., queries to find relationships between two or more variables; Leemkuil, 2006). This group of supports helps learners pause gameplay for a moment, analyze their gaming answers/solutions, and use organizational and integrational cognitive processes to learn the underlying concepts within the game.

The second type of support is modeling. This type of support provides an explication or illustration of how to solve a problem or perform a task in the game. The two most common supports categorized under the modeling category are: (1) scaffolding (Barzilai & Blau, 2014), and (2) worked examples (or expert solutions; Lang & O’Neil, 2008). Modeling can be provided either inside or outside of the game, by a peer, an expert, or the game itself; and it can be delivered verbally, graphically, or via animated form. One possible criticism regarding the inclusion of worked examples in a game is that learners can see a solution and then replicate it without actually thinking about the underlying concepts being used to solve the problem. However, with a good reward/penalty system in place, negative effects of using worked examples can be minimized. Also, providing partially worked examples can reduce the potential negative effect of fully worked examples. This is described in more detail later in this chapter where we present an example of integrating such worked examples in PP.

The third type of support is advice (e.g., Leutner, 1993), intended to guide the learner in the right direction without revealing the solution. Varied types of pre- or post-action advice (contextualized, adaptive or not) that are game-generated can be grouped under this category. For example, a hint can provide the learner with suggestions about what to do next in the game or provide an elaborated explanation about possible consequences of his/her action. Advice can consist of a short message asking the player to focus on a particular aspect of the task or give a cue about where to start.

The fourth type of support is collaboration (van der Meij, Albers, & Leemkuil, 2011), which may involve game talk with other players on a particular level or a gameplay strategy. Collaboration can help novice players figure out ambiguities in the game and better understand the knowledge and skills they need to learn. Many games allow for live chat and exchange of information among players. Alternatively, collaborative gameplay may be done with learners playing the game in dyads or small groups, then they can get involved in after-game discussions in online forums or in physical environments (e.g., a classroom).

The fifth learning support type is interactivity. This type of support focuses on soliciting active input from the learners when they process a learning support. Any type of learning support which is responsive to learners’ actions can be categorized under this group. For example, Moreno and Mayer (2005) designed their agent-based multimedia game with interactivity where students had to select roots, stems, and leaves that best helped plants survive on the planet. Another group of students used a different version of the game (i.e., with no interactivity). They interacted with a pedagogical agent who simply showed them pertinent information regarding the plants. The authors found that interactivity helped students learn and retain knowledge.
Narrative elements comprise the sixth type of learning support, where content can be integrated into the storyline of a game via narratives that contain surprises, foreshadowing, and fantasies. The narrative of a game provides a cognitive framework for the learners with which they can better learn and remember the underlying concepts in the game (e.g., Adams, Mayer, MacNamara, Koenig, & Wainess, 2012). This type of support can be seen, as Prensky (2001) pointed out, in genres such as adventure games or role-playing games.

The seventh type of learning support—and likely the most frequently used one—is feedback, especially formative feedback which is essential for learning (Shute, 2008). Given the high degree of interactivity existing in most games, feedback becomes critically important. As Shute (2008) notes, there are many types of feedback, but the two most common types used in educational games are corrective feedback (e.g., showing if an answer/solution is correct or not), and explanatory feedback (e.g., describing why the answer/solution was right or wrong). Cameron and Dwyer (2005) found statistically significant differences on all learning outcomes when feedback was included in the game versus when it was not.

Finally, modality (Ginns, 2005; Moreno & Mayer, 2002; Ritterfeld, Shen, Wang, Nocera, & Wong, 2009) refers to the representation of the support (e.g., auditory, visual, textual) and each type of modality can positively or negatively affect learning. For example, Moreno and Mayer (2002) found that learners remembered more of the materials, achieved better transfer, and rated more favorably virtual reality environments that used speech rather than on-screen text to deliver learning materials. Also, Ritterfeld and colleagues (2009) point out that multimodality is one of the most important facets of educational game success—providing learners with materials via different channels. Results of their study showed that multimodality positively affects knowledge gains for both short-term (at the posttest) and long-term (follow-up test) outcomes.

After conducting a moderator analysis, Wouters and van Oostendorp (2013), found out that among the 29 studies they examined, reflection, modeling, collaboration, modality, and feedback enhanced learning, but advice, interactivity, and narrative did not. Thus, the first group of support types is where we focus our attention. It should be noted, though, that the effectiveness of learning supports depends on how they are integrated into educational games.

Game-based Stealth Assessment for Adaptive Support of Learning

The underlying mechanism of adaptive learning support is real-time assessment and tracking of learners’ competency development during gameplay. Again, the challenge is validly and reliably measuring learning in games without disrupting engagement and leveraging that information to bolster learning. Our solution involves using stealth assessment (Shute, 2011) for crafting valid game-based assessments and dynamically linking those assessments to various learning supports. This methodology can contribute to the design of next-generation learning games that successfully blur the distinction between assessment and learning.

Stealth Assessment

For the past decade, we have been researching various ways to embed valid assessments directly into games with a technology called stealth assessment (see Shute, 2011; Shute & Ke, 2012; Shute & Ventura, 2013; Shute, Ventura, Bauer, & Zapata-Rivera, 2009). In 2011, we received funding from the Bill & Melinda Gates Foundation to build and test the stealth assessment technology in a game
we developed called *Physics Playground (PP)*. Stealth assessment is based on an assessment design framework called evidence-centered design (ECD; Mislevy, Steinberg, & Almond, 2003). In general, the main purpose of any assessment is to collect information that will allow the assessor to make valid inferences about what people know, can do, and to what degree (collectively referred to as “competencies” in this chapter). ECD defines a framework that consists of several conceptual and computational models that work in concert. The framework requires an assessor to: (a) define the claims to be made about learners’ competencies, (b) establish what constitutes valid evidence of a claim, (c) determine the nature and form of tasks that will elicit that evidence, and (d) determine how much evidence is required to support each claim.

Stealth assessment complements ECD by determining specific gameplay behaviors (specified in the evidence model) and linking them to the competency model (Shute & Ventura, 2013). As students interact with tasks/problems in a game during the solution process, they are providing a continuous stream of data (captured in a log file) that is analyzed by the evidence identification (EI) process. The results of this analysis are data (e.g., scores, tallies) that are passed to the evidence accumulation (EA) process, which statistically updates the claims about relevant competencies in the student model—the student’s individual copy of the competency model. The ECD approach combined with stealth assessment provides a framework for developing assessment tasks that are explicitly linked to claims about personal competencies via an evidentiary chain (i.e., valid arguments that serve to connect task performance to competency estimates) and are thus valid for their intended purposes. The estimates of competency levels can also be used diagnostically and formatively to provide adaptively selected levels, feedback, and other forms of learning support to students as they continue to engage in gameplay. *PP* uses these tools as the basis for developing adaptive learning supports. Again, the tricky part is embedding learning supports deeply in the game and not disrupting engagement.

The Original Version of Physics Playground—PPv1

Since its inception, *PP* has gone through various improvements regarding game design and embedding learning supports in the game. To clarify which version of *PP* we are referring to, we call the first version of *PP* without learning supports as PPv1, and the current version of *PP* with learning supports as PPv2. Also, we will use PPv2.1, 2.2, and 2.3 to refer to the usability studies we conducted using PPv2. Note that when the specifications we refer to are included in both versions, we simply use *PP*.

*PPv1* is nonlinear: players can choose any level in the game to play or replay. There is only one level-type in PPv1—sketching levels. The goal of all levels (or problems) in *PP* is to guide a green ball to hit a red balloon. Using the mouse or stylus, players draw objects on the screen, which “come to life” as physical objects when the mouse button/stylus is released. These objects interact with the game environment according to Newtonian mechanics and can be used to move the ball. When objects interact within the game environment, they act as “agents of force and motion” or just “agents”—simple machines in formal physics: ramp, lever, pendulum, and springboard. Figures 1 and 2 show screen shots from *PP* illustrating solutions using different agents and their associated physics concepts.
We used ECD to design PPv1 (Shute & Ventura, 2013). That is, we first established a simple physics competency model that included Newton’s laws of force and motion with two main facets, angular momentum and energy, that were associated with various agents (e.g., understanding potential and kinetic energy was statistically linked to ramp and springboard solutions). Second, evidence was defined as the behaviors demonstrated by a player in the game that would provide information about particular competency levels. For instance, the degree of understanding angular momentum is partly informed by evidence relating to the number of successful solutions involving pendulums. Third, task models provided a blueprint for creating all of the levels in PPv1, where each level focused on eliciting different agents (related to different physics principles) for solution. Levels also varied by difficulty. The difficulty of a problem was based on a number of factors, including: relative location of the ball to the balloon, number of obstacles present, number of agents required to solve the problem, and novelty of the problem. Difficult problems provide greater weight of evidence to the estimate of a competency level than easy problems. Also, “elegant” solutions (i.e., those using a minimal number of objects in the solution, suggesting mastery) give greater weight to competency level inferences than regular solutions.

Advancing Understanding of Game-based Learning

**Preliminary empirical findings.** We’ve now tested PPv1 in several studies to (a) validate the in-game (stealth assessment) measures, and (b) examine any learning of qualitative physics that may have occurred from playing the game, and the results are consistent. For instance, in Shute, Ventura, and Kim (2013), we found that performance data in PP as captured in the log files (e.g., use of a particular agent) significantly correlated with our external test scores, serving to validate the stealth...
assessment measures. In addition, we found that students (grades 8 and 9, n = 168) do improve in their qualitative physics understanding (t(154) = 2.12, p < .05) after four hours of gameplay with no content instruction or any other learning support. The pretest and posttest each consisted of 12 matched multiple-choice items related to relevant physics principles, similar to those in the Force Concept Inventory (Hestenes, Wells, & Swackhamer, 1992), which is a multiple-choice test designed to monitor students’ understanding of force and related kinematics. Finally, males and females demonstrated comparable learning gains and equally enjoyed playing PPv1 after controlling for pretest knowledge. The findings have been replicated (e.g., Shute et al., 2015), suggesting that this game has potential to foster motivation and learning in physics—for diverse learners.

**Fostering conceptual physics understanding.** Given these preliminary results, we wanted to substantially bolster learning in PP—from qualitative construal of physics to a deeper, more conceptual and formal understanding—via engaging and effective in-game learning supports that foster conceptual processing during physics-governed puzzle solving. In support of this approach, Hatano asserts that conceptual knowledge gives “meaning to each step of the skill and provide[s] criteria for selection among alternative possibilities for each step within the procedures” (1982, p. 15). Without this pairing between concepts and procedures, children develop only routine expertise: the ability to solve narrowly-defined, predictable, and often artificial (school-based) problems. Routine expertise is not very helpful outside of the school setting because it cannot be adjusted for and/or applied to real-life or unexpected situations.

Another reason to lay a solid conceptual physics foundation in PP is because even college students with acceptable grades in one or more physics courses have limited understanding of conceptual physics and hold erroneous views about the basic physical principles that govern the motion of objects in the world (Halloun, 1996; Reiner, Proffit, & Salthouse, 2005; Swann, 1950). Recognition of this problem has led to interest in the mechanisms by which physics students make the transition from informal (or naïve) 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—including simple machines like ramps, levers, and pendulums (devices designed to either change the magnitude or the direction of a force in PP), which are widely used to introduce physics concepts (Hewitt, 2009). Research on science education also indicates that learners’ hands-on experience with such machines (virtually and physically) supports understanding of physics concepts (Hake, 1998).

The enhanced game helps players lay the conceptual foundation of physics, before adding formalizations. In other words, we want PP players to learn physics at a deep and meaningful level before being introduced to formal physics terms and equations. For example, in PPv2, players first experience the relationship between force, mass, and acceleration when solving problems in the game (e.g., creating successful pendulum solutions) and then are introduced to relevant formalizations (e.g., F = ma, or Newton’s second law) via embedded learning supports.

**Core physics concepts to be assessed and supported.** PP behaves according to Newtonian (real-world) physics principles and dynamically responds to players’ interactions with the environment. To accomplish this responsiveness, PP performs a detailed formal simulation of a virtual physics “world” using actual, accurate physics formulas and calculations to account for mass, gravity, friction, momentum, and other physics concepts. Some of the physics concepts that we are assessing and supporting in the current research include Newton’s three laws of force and motion;
potential and kinetic energy; torque; collisions and conservation of linear momentum; and energy and dissipative forces.

**Challenges for games as formal learning tools.** Well-designed games typically possess the following game features: interactivity, ongoing feedback, and adaptive challenges. What good game designers have been doing intuitively for decades (i.e., applying these features within games as embedded learning support) is what we want to identify, codify, apply, and test in PP with the goal of creating a methodology that can be used in the design of next-generation learning games. Clark and his colleagues (e.g., Clark et al., 2011; Martinez-Garza, Clark, Nelson, Slack, & D’Angelo, 2013) have been tackling a similar problem of how to combine gameplay activities (in a game called Surge) with formal physics representations and terminology to support the learning of Newtonian mechanics. However, Clark and his colleagues (2011) found no significant gains for U.S. students on physics concepts after gameplay. They also reported that the in-game performance measures (e.g., number of replays, scores) were not related to students’ performance on physics tests. They concluded that more research is needed that provides “supports for students to help them articulate their intuitive understandings from game play with the explicit formal concepts and representations of the discipline” (p. 2192). Developing and testing these types of supports comprises our current research; specifically, technology innovations that include coupling stealth assessment processes with adaptive sequencing of levels along with just-in-time feedback—the latter via multiple representations, such as text, overlays, and animations.

**In-game Learning Supports in Physics Playground**

While different types of learning supports tend to promote learning across educational games (Wouters & van Oostendorp, 2013), details about particular features and their associated effectiveness of different types of learning supports are lacking in the literature (Johnson, Bailey, & Van Buskirk, 2017; Ke, 2016). Ke and Shute (2015) pointed out that the next generation of educational games will likely embody two related functions: (1) game-based stealth assessment, and (2) adaptive learning supports, which are based on the results of the in-game assessment. Effectively integrating the assessment and associated supports must rely on an iterative design and testing process.

In this section, we describe some of our processes related to iteratively developing, implementing, and testing various learning supports in PPv2.

**New Version of Physics Playground—PPv2**

Over the past two years, we have been designing and testing the effectiveness of a variety of learning supports in PPv1 to foster deep, more formal understanding of Newtonian Physics, without disrupting flow. We are finalizing the cognitive supports and working towards developing an adaptive, stealth assessment-based level selection algorithm. In this section, we describe the steps we took to design and develop the supports.

**Expanded competency model.** To expand the content to be measured and supported in the game, our first step was to revise PPv1 to extend the physics competency model. Using the Next Generation Science Standards (NGSS) as our guidepost, we worked with our two physics experts to select primary physics competencies and sub-competencies to be assessed in PPv2. We also identified all salient game behaviors (or observables) that can provide evidence of the proficiency status of
each variable in the competency model. After many revisions (Almond, Tingir, Lu, Sun, & Rahimi, 2017), we finally came up with the competency model shown in Figure 3. The model involves four primary competencies: force and motion, linear momentum, energy, and torque. The model serves as the foundation for subsequent design phases.

The model also serves as a foundation for the psychometric model driving the stealth assessment and adaptivity. To that end, the model must capture both hierarchical dependencies among the concepts and correlations in their acquisition, in the target community. To ensure that the final model captured these, multiple models were created by the psychometric team and presented to the physics experts, allowing them to weigh in on design alternatives (Almond, et al., 2017).

**New task types and levels.** The next step involved designing new task types that can elicit evidence of this more elaborated set of physics concepts. This resulted in the design of our new manipulation task type. Manipulation tasks require players to move three sliders (i.e., gravity, mass, and air resistance) and/or add external forces (i.e., static/dynamic blowers and puffers) to solve a level without drawing new objects. For instance, solving the *Whale* level (see Figure 4) requires players to adjust air resistance. In gameplay, increasing air resistance will slow down the falling of the ball and allow it to hit the balloon inside the whale’s mouth.
New learning supports. Across the past two years, we developed eight different learning supports for PPv2: (1) worked examples, (2) animations, (3) interactive definitions, (4) formulas, (5) Hewitt videos, (6) glossary, (7) hints, and (8) interactive tutorials.

In line with Wouters and van Oostendorp’s (2013) categorization, our worked examples serve the function of Modeling; our hints focus on Advice; and our animations, formulas, Hewitt videos, and glossary promote conceptual understanding via dynamic Modalities (i.e., each physics concept in the game can be presented across multimodal representations of the targeted physics knowledge). We selected Modeling, Modality, and Hints as the main types of support to include in the game because they are found as the most effective supports to elevate student learning relative to other learning supports (Wouters & van Oostendorp, 2013).

Usability Studies of Learning Supports

Three usability studies were conducted over the course of development of PPv2. The first version of PPv2 that we piloted—PPv2.1—consisted of 30 sketching levels and 30 manipulation levels, and the learning supports included worked examples, physics facts, advice, and Hewitt videos. The first usability study focused on gathering qualitative data on student opinions of both the task types and learning supports.

Usability Study 1—PPv2.1

Based on our observations and interviews, students enjoyed playing both sketching and manipulation levels. For the sketching levels, students enjoyed drawing on the screen and inventing creative solutions. However, sketching levels were reported as more difficult than manipulation levels by students. For the manipulation levels, students liked the direct maneuvering of the physics variables and the ability to see immediate results of the change in variables. They also liked that they were not limited by their ability to accurately draw and could focus more on controlling the movement of the ball.
Of all the learning supports included in the game, students preferred the worked examples. Worked examples are videos of expert solutions. All worked examples are less than a minute long with the majority being less than 30 seconds. The worked examples, created for 130+ game levels, can be viewed on our YouTube channel (https://www.youtube.com/channel/UCJpWi45D51ITxaj_NaClqjQ). Physics Facts was the least favorite support. It was a document containing definitions and examples of the relevant physics terms. Students reported it as an intensive reading, that it lacked visuals and/or interactions, and that it was not game-like. Another support feature (called Advice) consisted of level-specific suggestions on problem-solving thinking. However, the advice support was not available until after five minutes of playing a level. Due to this delay, many students did not access this support. The majority of students who did see the advice support reported it as non-helpful, vague, or confusing. Hewitt videos are an engaging series of cartoon videos explaining various physics concepts, developed by Paul Hewitt. The physics experts helped us select the most relevant videos for the game. With the permission of the author (Paul Hewitt), the team edited the length of each video to make it illustrate a targeted competency. Most students reported Hewitt videos as helpful.

Another major finding of the first usability study is that students never accessed the supports voluntarily even though the learning support is ever-present and accessible at players’ control. The learning supports were housed in a sidebar that opened from the left-side of the screen. The green triangular tab remained on the screen throughout every level, so students could click to open the supports any time during game play.

**Usability Study 2—PPv2.2**

The results of the first usability study led to revisions and additions to the learning supports in the game. In the second version of the learning supports, students access support through a help button in the lower-right corner of the screen (note: currently accessing supports is controlled by the player but in upcoming studies, we will examine the effects of player- vs. game-control of the supports). Clicking the help button triggered a pop-up window showing two options: “Show me the Physics” and “Show me a Solution”. The two options provide students two paths: learning support or gameplay support. “Show me the Physics” comprises the Modality-related, content-rich learning supports – where students can learn about physics phenomena via multiple representations (i.e., physics animations, interactive definitions, formulas, Hewitt videos, and a glossary). “Show me a Solution” focuses on game action-oriented, problem solution modeling (i.e., worked examples). The only revision made to the worked examples after the first usability study was to remove the audio track containing the physics explanations.

*Show me the Physics* leads the student to the physics support page showing the following options: “Animation,” “Definition,” “Formula,” Hewitt video”, and “Glossary” (note that the formula option is not present if the concept doesn’t have an associated formula or equation).

- **Animations.** We developed short (5 – 15 seconds) animations to illustrate the targeted physics concepts (e.g., gravity) in the game context. The team storyboards all the animations which were reviewed by the physics experts. The animations can be viewed [here](https://www.youtube.com/channel/UCJpWi45D51ITxaj_NaClqjQ) on our YouTube channel.

- **Interactive Definitions.** Originally the physics facts support, we turned this static element into an interactive, drag-and-drop quiz. Clicking on the definition option opens a
window showing an incomplete definition with five blanks to fill, five phrases to use, and relevant animation of the term/concept. Students watch the animation and drag each of the five phrases to the correct blanks within the definition. If the dragged phrase is not correct, it snaps back to its original place. When the blanks are correctly filled, a congratulation message pops up and displays the complete definition.

- **Formulas.** In collaboration with the physics experts, we created annotated mathematical formulas for the physics terms. Clicking on the formula option reveals the formula, along with a short explanation of each component/variable.

- **Hewitt Videos.** This is the same support used in the first version discussed earlier.

- **Glossary.** The glossary provides brief explanations of 28 physics terms. The terms have been selected, edited, and revised by the physics experts.

The second usability study focused on testing the effectiveness of the updated learning supports and the overall learning gains from gameplay and supports. Forty-four eighth-grade science students (23 males and 21 females) participated in a four-day quasi-experimental study. Half of the participants were selected to play the with-support version of the game while the others played the no-support version of the game. Students were provided with 60 levels in total (i.e., 30 sketching levels and 30 manipulation levels targeting nine physics concepts). Their conceptual physics understanding was measured by two parallel 18-item far transfer physics tests. Participants’ experiences and attitudes were also collected through a post-study survey.

Although previous empirical studies have shown playing PP improves students’ conceptual physics understanding (Shute, Ventura, & Kim, 2013), this specific usability study failed to detect any learning gains ($t(43) = .96, p = .37, d = 0.14$). In addition, the group playing without supports had higher posttest scores when holding pretest scores constant ($F(1,43) = 4.06, p = 0.05, d = 0.61$). Potential reasons for the results are that the gameplay time was insufficient, the incentives for support usage were not implemented, and the supports were not all effective. In spite of a four-day study, the actual gameplay time was less than two hours due to technical and logistical issues. Again, students favored worked examples. Viewing solutions might have helped students solve the game levels but not assisted the development of physics understanding. Adding an incentive system for learning support usage will help to limit the abuse of worked examples and direct more attention to the other supports that are intended to enhance physics understanding. Finally, the usability of certain in-game supports is not fully aligned with the gameplay flow. For example, tutorial levels were too long to be processed and internalized. Players had to replay the tutorial levels, which broke their flow of gameplay.

The implications of the second usability study led to further revisions of the learning supports. In the third version of the learning supports, when students click on the help button, the pop-up window shows three options: “Show me the Physics,” “Show me a Solution or a Hint,” and “Show me Game Tips”. The three options still provide two different kinds of supports: learning and gameplay. “Show me the Physics” stayed the same and contains animations, interactive definitions, formulas, Hewitt videos, and a glossary to aid in students’ development of conceptual physics. “Show me a Solution or a Hint” and “Show me Game Tips” are revised or new additions with a focus on game actions, solution modeling, and level-specific advice. Within these options students can access tutorial and game mechanic reminders and learn about “My
Backpack” which depicts their gameplay progress and allows them to customize the game environment.

We purposefully designed an incentive system for PPv2 to motivate usage of the physics supports. Now, clicking on *Show me a Solution or a Hint* activates a pop-up window with two buttons: “Show me a hint $0” and “Show me a solution $60.” If a student selects the former, a hint message pops up. Again, hints are free, partially-worked examples. Research shows that a partially-worked example or level-specific hint can facilitate learning (ter Vrugte, de Jong, Vandereruysse, Wouters, & van Oostendorp, 2017). Also, based on feedback from the previous two usability studies, we designed hints to help those who are struggling but are reluctant to watch full solutions. For instance, if a sketching level can only be solved by a springboard, the level-specific hint may be: “Try drawing a springboard.” The worked examples did not change. If a student elects to view a worked example, he or she will watch one of our video-recorded expert solutions after paying $60 as a disincentive. If the student does not have enough money, a message pops up telling them they have insufficient funds to access the solution and he/she needs to collect some gold/silver coins by completing more game levels and come back again.

Finally, *Show me Game Tips* is where students can review game mechanics, a visualized game tutorial, and learn about “My Backpack.” Clicking on the button leads to a page containing several navigation tabs. “Controls,” “Simple Machines,” and “My Backpack” tabs appear in sketching levels, and “Tools” and “My Backpack” appear in manipulation levels.

- **Controls.** When a student clicks this tab, a scrollable page appears showing game mechanics (e.g., nudge, draw an object, and delete an object) for sketching levels. The page shows the rules and explanations along with associated static images.

- **Simple Machines.** When a student clicks this tab, images of the four simple machines (i.e., lever, pendulum, ramp, and springboard) show up. This helps students remember how the agents work without having to go through the full tutorials again.

- **Tools.** Clicking on this tab provides a review of the rules for the sliders in manipulation tasks and a short explanation about other tools available (i.e., puffers and blowers). For example, students see the images of mass sliders and the bounciness function on the left-hand side of the page, and the corresponding text on the right-hand side (e.g., “Mass and bounciness only affect the ball”).

- **My Backpack.** This tab appears in both sketching and manipulation levels. See Figure 5. On the left the physics tab is open. Here students can see their progress in the game as well as estimates of their current level of physics understanding (soon to be linked with the real-time stealth assessment). On the right, the store tab is open. Here, they can change the type of the ball, background music, and background image. This is an additional component of the incentive system, as the customizations must be purchased by the students.
Another major revision coming out of the usability study using *PPv2.2* was the redesign of the game tutorials as sandbox game levels. In the first two usability studies, the tutorials were interactive videos. Students watched how to do something and then got the opportunity to try it. Observational data revealed that students were not retaining the information in the tutorial videos and students reported the tutorials were too long. Therefore, the tutorials in *PPv2.2* were changed to interactive levels with on-screen instructions and were used in *PPv2.3*—explained next. Sketching tutorial levels show how to draw simple machines. Manipulation tutorial levels show how to use the puffer or blower (that can exert a one-time and small force or a constant force, respectively), sliders (i.e., that control the values of mass, gravity, and air resistance), and the bounciness function. Students can access them either from the playgrounds or can view static images of the tutorial levels in “*Show me Game Tips*” button.

A final revision coming from the usability studies was the decision to develop a new type of learning support. The new support better integrates the support for learning and gameplay. We reviewed all the game levels, both sketching and manipulation, focusing on how the level was solved and the competencies with which it was linked. For each intersection of solution agent and competency, a video has been (or is being) developed. The videos illustrate the physics concepts through a worked example of the solution agent’s tutorial level (e.g., ramp, lever, pendulum). Narration and on-screen text with video pauses provide an overlay of the physics involved in the solution. These physics-related videos follow the same structure: (1) introduce the concept that will be presented in the video (e.g., “Here you are going to see how energy is transferred to a ball using a pendulum”), (2) state the concept (e.g., “gravitational potential energy is the energy of height…”), (3) demonstrate a failed attempt to solve a level in PP environment (i.e., the pendulum does not have enough angular height), and then (4) show a successful attempt to solve that level.

**Usability Study 3—PPv2.3**

The third usability study was conducted to investigate the effectiveness of the new learning supports (i.e., seven animations explaining the underlying physics concepts with text overlays and narration) when combined with game play—using *PPv2.3*. For the study, we chose two of the nine competencies with less overlap in our competency model: energy can transfer (ECT) and properties of torque (POT). Then, we selected a mixture of 30 sketching levels ranging from easy to hard with POT or ECT as their primary underlying physics concept. We also included the new set of sketching tutorial levels. So, in total, students had 35 levels to complete.
Our sample included 14 students (6 seventh graders, 8 eighth graders; 6 females, 8 males) from a small charter school in Florida. The study took place on one day for two hours. In the first 20 minutes, students completed a demographic questionnaire and pretest. Then, all the students played the game for 75 minutes in two stages: (1) the first 20 minutes: getting familiar with the game through the tutorials and freely accessing all the learning supports, and (2) the next 45 minutes: playing the game and accessing only the “physics supports” (in this stage the researchers prompted the students to access the “physics supports” after playing 3 levels or approximately every 8 minutes). At the end of the gameplay, students completed the posttest, and the game and learning support satisfaction questionnaires (all the tests were administered online using Qualtrics).

Despite the limitations of this usability study (small sample size and short gameplay time), we obtained some useful findings that can help us improve the game for future studies. Examining pretest and posttest scores, students scored significantly higher on the posttest compared to the pretest ($M_{pre} = 0.56$, $M_{post} = 0.63$, $t(13) = -2.20$, $p < 0.05$, Cohen’s $d = 0.60$). Also, the analysis of students’ overall game and learning supports satisfaction (ranging from 1—Strongly Disagree, to 5—Strongly Agree) showed that students enjoyed playing the game ($M = 4.24$, $SD = 0.62$), and they saw the learning supports as useful and easy to use ($M = 3.99$, $SD = 0.51$). Moreover, males and females equally enjoyed the game and the supports. These findings suggest that we are on the right path, and we will continue to improve the game for the future studies. The reflection on students’ learning experience also prepares us for the next phase of the project—implementing an adaptive algorithm into the game.

**Measurement of competencies in Physics Playground**

To make real-time estimates of students’ competency states, our evidence model has two processes: (a) evidence identification (EI) – the scoring of student actions, called “observables,” recorded in the log file, and (b) evidence accumulation (EA) – i.e., making inferences about competency states given the observables. The EA process for PPv2 is implemented using Bayesian networks based on an evidenced-centered assessment design framework (Almond et al., 2015). At the hub is a central Bayesian network representing the competency model, and the spokes are evidence models—Bayesian network fragments associated with each game level. As the player interacts with the game, the game engine logs key events to an event database.

The scoring engine works as follows: (1) Each user is given a student model—an individual copy of the competency model—in which information about that student is stored. (2) When information comes from the game engine that a student has completed a particular level, the evidence identification (EI) engine is run to extract the key observable features from that student’s performance according to the specifications in the evidence model. (3) The Evidence Accumulation (EA) engine attaches Bayesian network fragments for the evidence model to the student model and updates the competency estimates for that student.

If PPv2 is being run in an adaptive mode, the activity selection process runs on top of the EA code. The adaptive algorithm consists of an outer loop and an inner loop. The outer loop follows the critiquing strategy (Barr & Feigenbaum, 1982) where a series of target competencies are chosen in sequence. In each case, tasks and activities related to the target competency are presented to the student until the probability of mastery exceeds either a high or low threshold. If the student passes the high threshold, they are passed on to the next competency in sequence. If they fall below
the low threshold, they are routed to remedial activities. In the inner loop, the expected weight of evidence (EWOE; Madigan & Almond, 1996) algorithm is used to select the task which provides the highest EWOE for the currently targeted proficiency. Shute, Hansen and Almond (2008) speculate that tasks with high EWOE are in the center of the zone of proximal development (Vygotsky, 1978) of the student.

Because the Bayesian network scoring model is closely tied to the evidence-centered design (ECD) models, it is necessary to be able to quickly update the Bayesian networks when the ECD models change, and vice versa. The R package Peanut (Almond, 2018) automates this translation process, providing both network and tabular views of the assessment. It also uses the DiBello parameterization for the Bayesian networks, allowing cognitive scientists to specify relationships in terms of difficulties and discriminations instead of conditional probabilities. The probabilities used in the network are initially specified by experts and then revised as data become available to calibrate the network (Kim, Almond & Shute, 2016).

The use of learning supports could be treated as additional game levels, with their own evidence models. Accessing a learning support would provide a small amount of positive evidence that the player has mastered the associated competency. We do not think that the availability of the learning supports will influence the measurement. That is, Shute, Hansen and Almond (2008) found that (a) students assigned to an elaborated feedback condition showed significantly greater learning gains than students who received accuracy-only feedback, and (b) those students in the elaborated feedback condition showed a higher (although not significantly so) correlation with posttest scores than the accuracy-only feedback condition. Evidence models for learning supports have not yet been implemented but are a possibility for future versions.

Discussion

In this chapter we described the general effectiveness of educational games with embedded learning supports, discussed various types of learning supports identified in the literature, and illustrated how we designed, developed, and have begun testing different learning supports in our educational game—PP. After conducting the first two usability studies, we found that students were not adequately motivated to access the more helpful learning supports (i.e., physics related supports) given the absence of an appropriate in-game reward system. Therefore, we revised the game and supports to (a) further clarify and enhance the appearance and interactivity of the learning supports, (b) provide easier, more direct access to the supports, and (c) set up a compelling and functional reward system.

Moving forward, there are a number of potential avenues for research in this area, such as determining the degree to which a reward system actually influences students’ play experience and motivation to access learning supports in the game. Towards that end, we are optimizing the learning supports and the game reward system. As the third usability study demonstrated, students had positive opinions of the new learning supports and showed significant learning gains in an abbreviated gameplay experience. We will continue developing the new support videos, following these results.

We are also developing affective supports to complement the cognitive supports that we have developed (not focused on in this chapter). Moreover, we have created a new “near transfer” physics test (where items have the look and feel of the game) to accompany our current “far
 transfers” test (where items are more formal and inspired by the Force Concept Inventory; Hestenes, Wells, & Swackhamer, 1992) to see the degree to which intuitive physics conceptualizations vs. formalizations are acquired from gameplay. Finally, we are using stealth assessment technology with an in-game adaptive algorithm to select the best next level for a person—one that is not too difficult nor too easy and related to the targeted physics concept.

This sampler of ongoing research will help us and other researchers in the field figure out ways to optimize the design and delivery of learning supports that may be unobtrusively incorporated into games. The process should be iterative and provide research-backed evidence on: (1) the effects of different types of cognitive and affective supports that promote formal learning and enjoyment in educational games; (2) the timing and control of such supports (e.g., when should they be available, and who—computer or player—controls the delivery; and (3) the factors that mediate the influence of supports on learning and gameplay.

On a broader note, this ongoing research involves developing and testing a specific methodology for crafting valid game-based assessments, and dynamically linking those assessments to various learning supports. This methodology will contribute to the design of next-generation learning games that successfully blur the distinction between assessment and learning. Additional contributions of this research to the field of learning sciences include: (a) the design heuristics and instructional effectiveness of an educational game designed using principles of instructional, game, and assessment design; (b) advancing our understanding of the implementation strategies and influences of different kinds of learning supports (e.g., visualizations and explanations) in the context of game-based learning; and (c) presenting analytical conjectures that can be incorporated into other types of STEM learning games, and linked with the Common Core Standards and Next Generation Science Standards.

Acknowledgements

We wish to express our gratitude to the funding by the US National Science Foundation (NSF #037988) and the US Department of Education (IES #039019) for generously supporting this research.

References


