Chapter 4
Game-Based Learning Analytics in Physics Playground

Check for updates

Valerie Shute, Seyedahmad Rahimi and Ginny Smith

1 **Abstract** Well-designed digital games hold promise as effective learning environ-
2 ments. However, designing games that support both learning and engagement with-
3 out disrupting flow [1] is quite tricky. In addition to including various game design
4 features (e.g., interactive problem solving, adaptive challenges, and player control
5 of gameplay) to engage players, the game needs ongoing assessment and support
6 of players' knowledge and skills. In this chapter, we (a) generally discuss various
7 types of learning supports and their influence on learning in educational games, (b)
8 describe stealth assessment in the context of the design and development of particular
9 supports within a game called Physics Playground [2], (c) present the results from
10 recent usability studies examining the effects of our new supports on learning, and
11 (d) provide insights into the future of game-based learning analytics in the form of
12 stealth assessment that will be used for adaptation.

1 Introduction

14 *Play is often talked about as if it were a relief from serious learning. But for children,*
15 *play is serious learning. —Fred Rogers*

16   As noted in the quote above, Mr. Rogers, along with many others before him, rec-
17 ognized the crucial link between play and learning. If true, then why are our schools   AQ1
18 more like factories than playgrounds? Before explaining this reality, first imagine
19 the following: Public schools that apply progressive methods—such as individu-
20 alizing instruction, motivating students relative to their interests, and developing
21 collaborative group projects—to achieve the goal of producing knowledgeable and
22 skilled lifelong learners. The teachers are happy, they work hard, and are valued

V. Shute (✉) · S. Rahimi · G. Smith
Florida State University, Tallahassee, USA
e-mail: vshute@fsu.edu

S. Rahimi
e-mail: sr13y@fsu.edu

G. Smith
e-mail: glsmith2@fsu.edu

by the community. In addition, they hold leadership roles in the school and work individually and collectively to figure out the best ways to reach and teach their students. These same teachers create new textbooks and conduct research to see whether their methods worked. School days are structured to allow teachers time to meet and discuss their findings with colleagues.

Is this an ideal vision of schools of the future? Yes and no. According to Ravitch [3], the image above describes several model public schools in the USA in the 1920s and 1930s, inspired by John Dewey's vision of education (e.g., the Lincoln School at Teachers College in New York, and the Winnetka, Illinois, public schools). These schools were engaging places for children to learn and were attractive places for teachers to teach; they avoided the monotonous routines of traditional schools [4].

So, what happened to these exciting experiments of educational reform, and more importantly, what lessons can we learn from them? First, according to Kliebard [5], they failed because the techniques and founding ideas were misapplied by so-called experts who believed that mass education could be accomplished cheaply, employing low-paid and poorly trained teachers who would either follow their manuals or stand aside while students pursued their interests. Second, they failed because the reforms rejected traditional subject-matter curricula and substituted vocational training for the 90% of the student population who, at the time, were not expected to seek or hold professional careers (see [6], "The Elimi—nation of Waste in Education"). Finally, this period also saw mass IQ testing (e.g., [7]) gaining a firm foothold in education, with systematic use of Terman's National Intelligence Test in senior and junior high schools. The testing was aimed specifically at efficiently assigning students into high, middle, or low educational tracks according to their supposedly innate mental abilities.

In general, there was a fundamental shift to practical education going on in the country during the early 1900s, countering "wasted time" in schools and abandoning the classics as useless and inefficient for the masses. Bobbitt, along with some other early educational researchers and administrators such as Ellwood and Ayers [5], inserted into the national educational discourse the metaphor of the school as a "factory." This metaphor has persisted to this day; yet if schools were actual factories, they would have been shut down years ago.

How can we counter this entrenched school-as-factory metaphor? One idea that has garnered a lot of interest lately is to use well-designed digital games as learning environments. Over the past couple of decades, research in game-based learning demonstrates educational games are generally effective learning tools (e.g., [8–10]). When people play well-designed games, they often lose track of time—i.e., experience the state of flow [1]. Teachers try to engage students with learning materials, but the engagement is usually not comparable to that experienced with good video games [10, 11]. Digital game-based learning can be defined as digital activities with goals, interaction, challenges, and feedback that are designed to integrate learning with gameplay.

There is no archetype for game-based learning. That is, games vary by content (e.g., level of narrative, subject matter), design (e.g., 2D, 3D, amount, and quality of graphics), genre (e.g., first-person shooter games, puzzles), and player configuration

(e.g., single player, multiplayer, competitive, and cooperative). The complicated part is designing games that support learning and engagement without disrupting flow [1]. For example, Habgood, Ainsworth, and Benford [11] suggest that when the learner is still figuring things out in the game (e.g., learning the basic game mechanics) providing learning content at that point is not a good idea.

Research on game-based learning also recommends the use of learning supports or scaffolds to aid in student knowledge and skill acquisition and transfer, specifically using a mixture of supports in the game, delivered via various modalities [12]. Players may need different types of learning support at different points during gameplay (e.g., more scaffolding at the beginning of the game) or they may prefer a different type of support (e.g., one might not want to see a solution, but instead just receive a hint). However, the research on learning supports and scaffolding used in learning environments in general is conflicted. Some researchers (e.g., [13]) note that learning environments that allow for full autonomy (i.e., student control), without explicit supports, can be more engaging and effective environments than those without such freedom. Clark, Tanner-Smith, and Killingsworth [8] concluded from their meta-analysis that extra instruction (after gameplay, in the form of learning support) did not produce any significant learning differences between game and non-game conditions where compared. However, Wouters and van Oostendorp [14] conducted a meta-analysis on the topic and, overall, found a positive, moderate effect of learning supports ($d = 0.34$, $z = 7.26$, and $p < 0.001$), suggesting the use of learning supports in games can, in fact, improve learning.

The challenge in the design of game-based learning is not just on how to integrate learning through various design features and supports, but also on how to accurately assess the player's knowledge and skills, in real time, and at an actionable grain size. The use of analytics, specifically, stealth assessment [15] built through evidence-centered design [16] is one possible solution. Evidence-centered design (ECD) is a framework to build valid assessments and generate estimates of student performance. It consists of conceptual and computational models working together. The three major models include the competency model, the evidence model, and the task model. The competency model is comprised of everything you want to measure during the assessment. The task model identifies the features of selected learning tasks needed to provide observable evidence about the targeted unobservable competencies. This is realized through the evidence model, which serves as the bridge between the competency model and the task model.

Stealth assessment is a specialized implementation of ECD, where assessment is embedded so deeply into the learning environment it is invisible to the learners [17]. Stealth assessment for game-based learning begins with a student immersed in gameplay, producing a myriad of performance data, all captured in the log file. Next, the automated stealth assessment machinery measures the observables from the logfile data. It then outputs the results of the analysis to the student model (i.e., an individualized competency model based on each student's data) which then provides estimates about the current state of the competencies for each individual student. These estimates are used to provide personalized feedback and other types of learning support to the player who continues to play the game and produce more performance

113 data. Thus, the stealth assessment provides real-time estimates as the cycle continues
114 (for more, see [17]).

115   The use of analytics in the form of stealth assessment has many benefits. In a
116 well-designed video game, with embedded stealth assessment, students are fully
117 engaged in the experience. Student performance during this level of engagement
118 enables more accurate extraction of students' knowledge and skills. Test anxiety can
119 cause students to perform below their actual ability on tests. Because it is designed
120 to be unobtrusive, stealth assessment frees students from the anxiety of traditional
121 tests and thus improves the reliability and validity of the assessment (e.g., [18, 19]).
122 Another benefit is that the stealth assessment can provide information about students'
123 competencies at a fine grain size. When compared with conventional assessments
124 like multiple-choice formats that yield a single summative score at the end, stealth
125 assessment delivers more valid, reliable, and cumulative information about a stu-
126 dent's knowledge and/or skill development. Its automation means teachers do not
127 need to spend time on tedious tasks such as calculating scores and deriving grades.
128 Finally, stealth assessment models, once developed and validated, can be recycled in
129 other learning or gaming environments through the adjustment of the evidence and
130 task models to the particular game indicators (e.g., [20]).

131   While stealth assessment can provide accurate, detailed information about student
132 performance, it can also provide adaptive support. For example, different types of
133 learning support can be employed and tailored, per student. That is, the what, how, and
134 when of learning supports can be fit to the current needs of individuals. Effectively
135 integrating the assessment and associated supports relies on an iterative design and
136 testing process, with an eye toward adaptivity—where the supports are available or
137 delivered at the right time, and in the right form to maximally enhance learning.
138 Figure 1 illustrates the flow of events in the game, based on information captured
139 in the log file, automatically scored, and accumulated via the stealth assessment's
140 models.

141   In our design-based research project, we aim to develop and test a methodology for
142 crafting valid and engaging game-based assessments and dynamically linking those
143 assessments to in-game learning supports (i.e., an adaptive algorithm and ongoing
144 feedback; see link 4 in Fig. 1). This methodology will contribute to the design of next-
145 generation learning games that successfully blur the distinction between assessment
146 and learning and harness the power of gameplay data analytics.

147   In this chapter, we (a) review the literature on various learning supports and
148 their influence on learning and performance in educational games, (b) describe our
149 own experiences with stealth assessment and the design and development of different
150 learning supports within a game called Physics Playground [2], (c) present the results
151 from recent usability studies examining the effects of our new supports on learning,
152 and (d) provide insights into the future of game-based learning analytics in the form
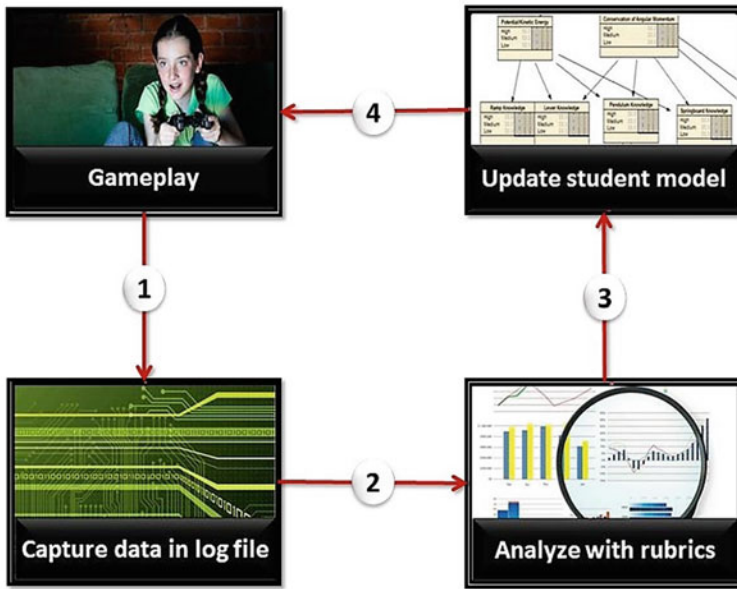153 of stealth assessment that can be used for adaptation.

Fig. 1 Stealth assessment cycle

2   Review of the Effects of Learning Supports in Games

Many kinds of learning supports have been used and tested in educational games and other kinds of learning environments. Overall, the results are mixed.

2.1   Types of Supports

In their meta-analysis, Wouters and van Oostendorp [14] identified 24 different types of learning supports and grouped them into ten categories. Here, we limit the focus to three of their categories of support: modeling, advice, and modality. The category of *modeling* includes supports that provide an explication or illustration of how to solve a problem or perform a task in the game. The two most common supports within the modeling category are: (1) scaffolding [21] and (2) worked examples (or expert solutions) [22]. The main purpose of scaffolding is to focus attention via the simplification of the problem at hand [23]. This can be accomplished by providing constraints to the problem that increase the odds of a learner's effective action as they focus attention on specific features of the task in an otherwise complex stimulus field. The main purpose of worked examples is to clearly demonstrate a solution to the task, via human or computer. One possible criticism of this category of support

170 is that learners can replicate a shown solution without having to think about the
171 concepts used to solve the problem.

172   The category of *advice* (e.g., [24]) refers to support that is intended to guide the
173 learner in the right direction without revealing the solution (as occurs with worked
174 examples). All types of advice (contextualized, adaptive or not) that are game-
175 generated can be grouped under this category. Many popular role-playing games
176 provide advice or hints through characters that players encounter in the game world.
177 These characters can give textual hints during dialogs with the player. Other games
178 allow players to buy hints with earned game rewards like coins or points. Generally,
179 including hints/advice in games is intended to provide support for struggling players,
180 but do they help learning? That likely depends on the type of hint provided (e.g.,
181 abstract vs. concrete), and how it is presented (e.g., [25]).

182   Finally, *modality* [26, 27] [12], like the name indicates, comprises learning sup-
183 ports provided via different modalities (e.g., auditory, visual, textual). Each type
184 can positively or negatively affect learning. For example, Moreno and Mayer [27]
185 found learners remembered more educational content, showed more transfer, and
186 rated more favorably virtual reality environments that used speech rather than on-
187 screen text to deliver learning materials. Providing materials via different channels, or
188 multimodality, is an important component of successful educational games [12]. Rit-
189 terfeld and colleagues found that multimodality positively affects knowledge gains
190 in both short-term (i.e., immediate posttest) and long-term (i.e., delayed posttest)
191 evaluations.

2.2 Timing of Supports

193 The two main questions about learning supports concern what to present (described
194 above), and when to make it available. Csikszentmihalyi [1] claimed that learners
195 learn best when they are fully engaged in some process—i.e., in the state of flow.
196 Inducing flow involves the provision of clear and unambiguous goals, challenging
197 yet achievable levels of difficulty, and immediate feedback (e.g., [28]). Based on flow
198 theory, a task that is too difficult can be frustrating and/or confusing while a task that
199 is too easy may be boring, thus the optimal state (of flow) resides between the two.
200 Similarly, Vygotsky's zone of proximal development (ZPD) suggests that learning is
201 at its best when the learning materials are just at the outer edges of students' existing
202 level of understanding and ability [29]. Considering these two aspects of deep learn-
203 ing—facilitating the state of flow and providing materials compatible with learners'
204 ZPDs—adaptive learning environments such as games can be used to facilitate both
205 by adapting to learners' current competency state(s).

206   In this section, we define adaptivity—related to the timing of supports—as the
207 ability of a device to alter its behavior according to changes in the environment.
208 In the context of instructional environments, adaptivity can help to provide person-
209 alized instruction for different learners and facilitate the state of flow throughout
210 the learning process. An adaptive learning environment should monitor various (and

often evolving) characteristics of learners then balance challenges and ability levels to improve learning (for more details on adaptivity in learning environments, see [30]).

One way to include adaptivity in educational games is to use micro-adaptation [31, 32]. This approach entails monitoring and interpreting the learner's particular behaviors, as with stealth assessment. Micro-adaptivity then may provide the learner with appropriate learning supports and/or adjust various aspects of the game (e.g., level difficulty) based on the student model estimates without disrupting the state of flow [31]. Adaptive games can adapt challenges to the current estimated levels of player's knowledge and skills [1], [29] and provide formative feedback [33] and other types of support in unobtrusive ways [34].

In summary, previous findings suggest that the content of the supports, as well as the timing of their availability/delivery, should be carefully designed according to the game features to achieve specific instructional purposes. Cognitive supports are needed in the game to bolster deep conceptual learning. In Physics Playground, this means helping students move from a qualitative, informal understanding of physics to a deeper, more conceptual, and formal understanding. In support of this approach, Hatano asserts that conceptual knowledge gives "meaning to each step of the skill and provides criteria for selection among alternative possibilities for each step within the procedures" ([35], p. 15). Without a pairing between concepts and procedures, students develop only routine expertise, which is the ability to solve a narrowly defined set of 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 (see [35, 36]).

We are interested in supporting adaptive expertise, which requires a student to develop conceptual understanding which, in turn, allows that student to invent new solutions to problems and even new procedures for solving problems. However, providing such support in games is more complicated than in other types of interactive learning environments. Cognitive support in games must reinforce emerging concepts and principles to deepen learning and engender transfer to other contexts, but without disrupting engagement while learners are immersed in gameplay.

We now present a case study illustrating how we have been incorporating and testing various supports in our game called Physics Playground.

3 *Physics Playground*—Evolution of Learning Supports

In this section, we elaborate on the process we have gone through to design, develop, test, and revise our learning game, *Physics Playground* (PP). From its inception, *PP* has gone through various changes which led to the development of different versions of the game. For simplicity, we refer to the first version of *PP* as *PPv1*, and to the current version of *PP* (with new task types, learning supports, an incentive system, open student model, and other features) as *PPv2*. Finally, if what we are referring to is general, we simply use the term *PP*.

3.1   The Original Physics Playground—PPv1

PP is a two-dimensional physics game designed to enhance physics understanding [2]. The goal in *PP* is simple—hit a red balloon using a green ball. *PPv1* includes only one type of game level: *sketching*. Using a mouse or stylus, players draw objects on the screen, create simple machines (i.e., ramp, lever, pendulum, or springboard), and target the red balloon with the green ball (see Fig. 2).

As shown in Fig. 2, the solution for the level called *Chocolate Factory* is a ramp affixed to the top part of the level using a pin and including an adequate slope which can guide the ball to the balloon.

In *PPv1*, we used stealth assessment technology [15] to measure player's conceptual understanding of physics related to: (1) Newton's laws of force and motion, (2) potential and kinetic energy, and (3) conservation of angular momentum [37]. Also, *PPv1* was used to measure non-cognitive competencies such as persistence [38] and creativity. Across multiple studies, we consistently found that (1) *PP* can foster motivation and improve learning and (2) the embedded stealth assessment measures are reliable and valid—significantly correlated with external measures (see [38]). Our primary goal, however, has always been improving physics understanding in a fun way—without disrupting flow. To that end, we took a step further to design and develop a new version of *PP* with a broader scope and adaptive learning supports.
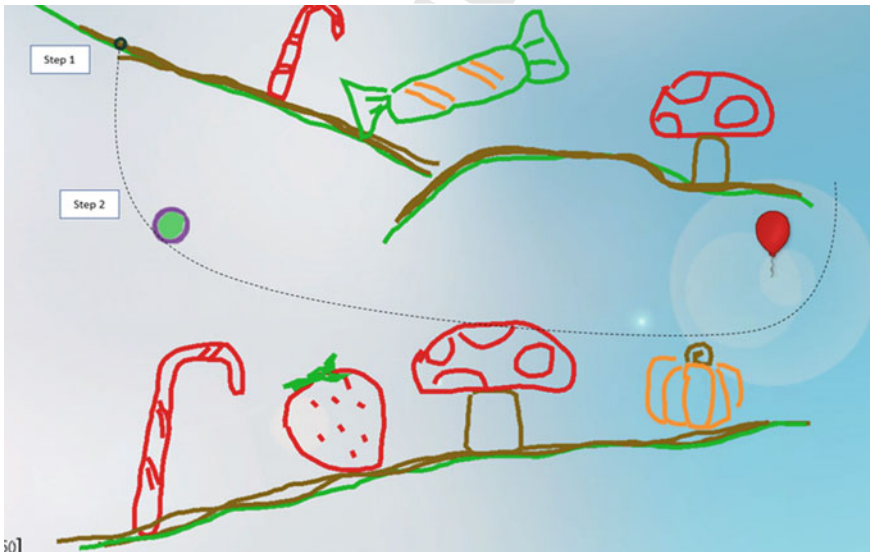

Fig. 2 *Chocolate Factory* level in *PPv1*

3.2 The Current Physics Playground—PPv2

The first step we took to develop *PPv2* was to redefine our previously rather sparse physics competency model. The new physics competency model (see Fig. 3) was guided by the Next Generation Science Standards (NGSS) and designed through an iterative process with the help of two physics experts.

New Task Type. The expanded competency model required the addition of new game tasks to the task model to elicit the new evidence. We needed to accurately measure students' proficiency levels per concept with the stealth assessment, so we designed a new task type, *manipulation levels.* In manipulation levels, drawing is disabled, and new features are used to move the ball to the balloon. The new features include (1) sliders related to mass, gravity, and air resistance, (2) the ability to make the ball bounce by clicking the bounciness checkbox, and (3) new sources of exerting external force (e.g., puffer, and static and dynamic blowers) to solve a level. For example, Fig. 4 shows a manipulation level called *Plum Blossom*. In a manipulation level, students get an initial situation with a predefined value for each slider. Then, students can manipulate the variables (i.e., sliders) to solve the level. When the *Plum Blossom* level is played initially, the ball falls, due to gravity, and it is not possible to elevate the ball and hit the balloon. To solve *Plum Blossom*, the player must change the gravity value to zero and use the blue puffer on the left side of the ball to exert a little force. With no gravity, the ball moves slowly to the right and hits the balloon. We designed and developed 55 new manipulation levels targeting various physics concepts in our physics understanding competency model.
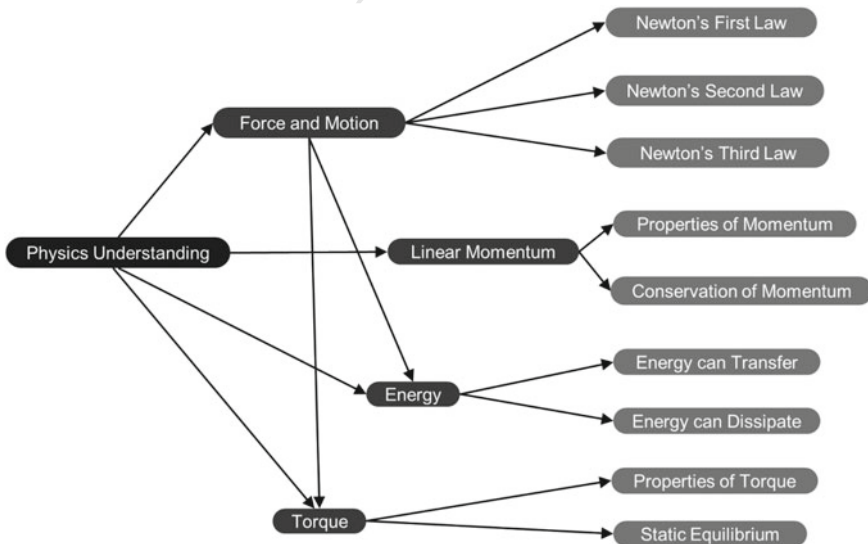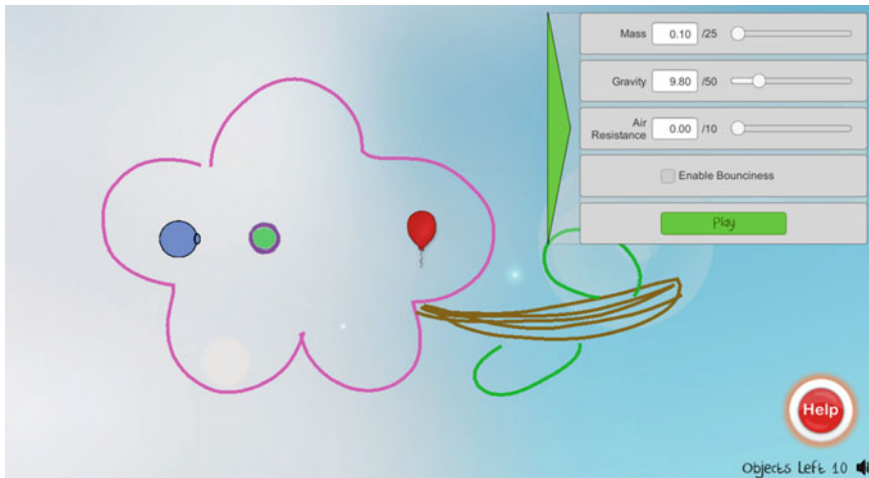

Fig. 3 Competency model for physics understanding in PPv2

Fig. 4 *Plum Blossom* level in *PPv2*

²⁹³ We tested the new task type in our first usability study. Based on our obser-
²⁹⁴ vations and interviews, students enjoyed playing both sketching and manipulation
²⁹⁵ levels. For the sketching levels, students enjoyed drawing on the screen and invent-
²⁹⁶ ing creative solutions. However, sketching levels were reported as more difficult than
²⁹⁷ manipulation levels by students. For the manipulation levels, students liked the direct
²⁹⁸ maneuvering of the physics variables and the ability to see immediate results of the
²⁹⁹ change in variables. They also liked that they were not limited by their ability to
³⁰⁰ accurately draw and could focus more on controlling the movement of the ball.

³⁰¹ Along with new task types, we also developed other features for the game, such as
³⁰² new game tutorials, the help interface and support content, and an incentive system.

³⁰³ **Game Tutorials**. Originally, the game tutorials were interactive videos, placed
³⁰⁴ in two separate playgrounds—sketching tutorials and manipulation tutorials. The
³⁰⁵ tutorials introduced essential game tools relevant to our two task types. Students
³⁰⁶ watched how to do something and then had an opportunity to try it. Usability testing
³⁰⁷ revealed that the tutorials were not particularly effective. They were too long, and
³⁰⁸ students could not accurately recall the information later when playing the game.
³⁰⁹ Based on these results and several rounds of revision, the tutorials are now interactive
³¹⁰ levels with on-screen instructions. Sketching tutorials illustrate how to draw simple
³¹¹ machines. For example, in Fig. 5, you can see the lever tutorial, with on-screen,
³¹² step-by-step instructions. If students follow the instructions, they can easily solve
³¹³ the level, get a silver coin ($10), and move to the next tutorial. Manipulation tutorials
³¹⁴ show how to use the puffer/blower (that can exert a one-time and small force or a
³¹⁵ constant force), sliders (i.e., for mass, gravity, and air resistance), and the bounciness
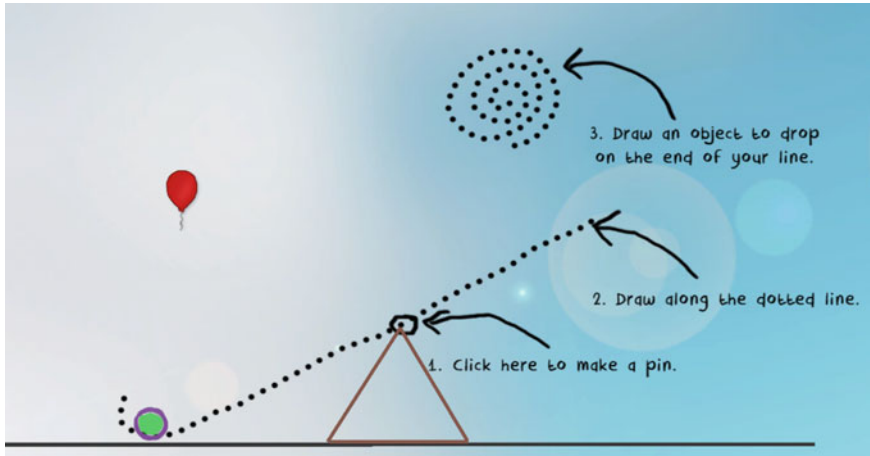³¹⁶ function.

Fig. 5 Lever tutorial level in *PPv2*

Learning Support. When designing the learning supports for *PP,* we had two major components to develop: (1) the location in the game and user interface for the supports and (2) the content and type of supports to offer.

Support Location and Interface. In the first version of *PPv2*, the learning supports were accessed via the "Support Kit" button located on the left side of the screen. Clicking on the button opened the support menu (Fig. 6). However, in the first usability study, students generally did *not* voluntarily click the button to open the

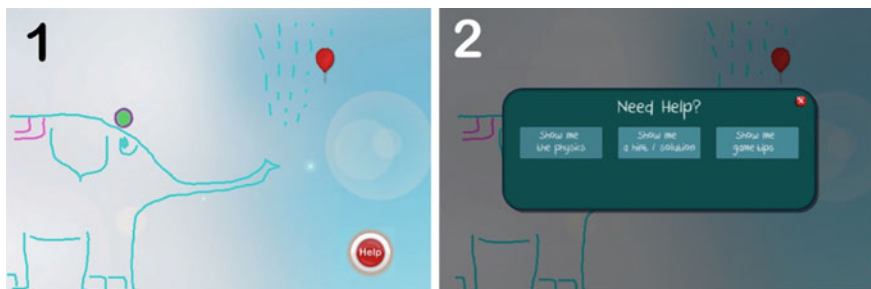**Fig. 6** Old support menu in *Flower Power* level in *PPv2*

Fig. 7 New "Help" button (left) and help menu (right) in *PPv2*

324 help menu. Consequently, we decided to revise the color, name, and position of the
325 button to make it clear and visually appealing. Thus, we designed a "Help" button.

326   The current support interface of the game begins in a level with a player clicking
327 the help button, located in the lower-right corner of the screen (Fig. 7). This trig-
328 gers a pop-up window showing three options: "Show me the physics", "Show me a
329 hint/solution," and "Show me game tips."

330   The first two options provide two different paths: learning support or gameplay
331 support. "Show me the physics" comprises the modality-related, content-rich learn-
332 ing supports where students can learn about physics phenomena via multiple rep-
333 resentations. "Show me a hint/solution" focuses on game action-oriented, problem
334 solution modeling. Finally, *Show me Game Tips* is where students find game rules and
335 tutorial reminders. Below are descriptions of each of these support options, including
336 their development process.

337   *Support Content*. In parallel with designing and developing the support interface,
338 we developed numerous learning supports for *PPv2*: (1) worked examples, (2) ani-
339 mations, (3) interactive definitions, (4) formulas, (5) Hewitt videos, (6) glossary,
340 and (7) hints. In line with Wouters and van Oostendorp's categorization [14], our
341 *worked examples* serve the function of modeling; our *hints* focus on advice; and our
342 *animations*, *formulas*, *Hewitt videos*, and *glossary* promote conceptual understand-
343 ing via dynamic modalities (i.e., each physics concept in the game can be presented
344 across multimodal representations of the targeted physics knowledge). We designed,
345 developed, tested, and revised these learning supports across three usability studies.
346 Each usability study focused on a different set of supports.

347   **Show me the Physics**. Clicking *Show me the Physics* leads the student to the
348 physics support page showing the following options: "Animation", "Definition,"
349 "Formula," Hewitt video," and "Glossary" (note that the formula option is not present
350 if the concept does not have an associated formula or equation, see Fig. 8).

351   *Animations.* The animations integrate gameplay and support for learning. The
352 team reviewed all the game levels, both sketching and manipulation, focusing on
353 how the level was solved and the competencies with which it was linked. A separate
354 animation has been or will be developed for each intersection of solution agent (i.e.,
355 simple machine) and competency. The new support videos utilize the game levels
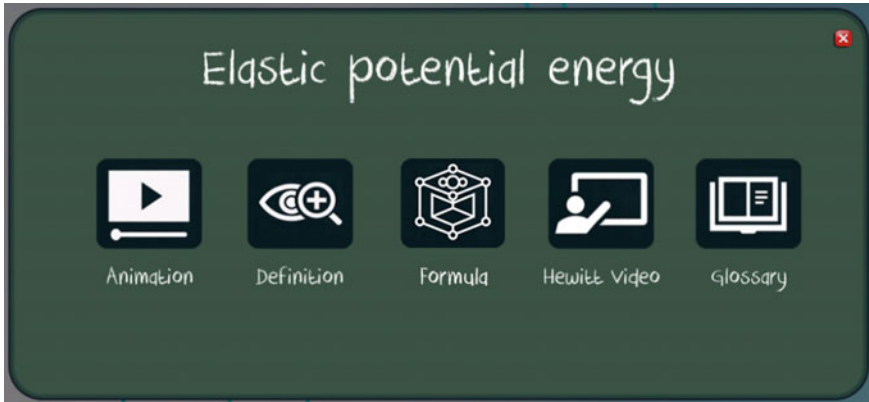
Fig. 8 "*Show me the Physics*" menu in *PPv2*

356 to illustrate the physics concepts through failed and successful solution attempts.
357 Narration and on-screen text with video pauses provide an overlay of the physics
358 involved. The new physics animations, with narration, connect the physics concepts
359 with how they are applied in the game to solve a level.

360   *Interactive Definitions.* Originally, this was an online document entitled, Physics
361 Facts, which when clicked, led to a non-interactive physics term list, showing def-
362 initions and short examples. The results of the first usability test showed students
363 did not like or use this support. They reported it as an intensive reading, that lacked
364 visuals and/or interactions, and was not at all like the other game components. Based
365 on these results, we transformed the boring, static support into an interactive, drag-
366 and-drop quiz. Players now, interactively, construct definitions of terms, like a Cloze
367 task [39]. Clicking definition opens a window showing an incomplete definition with
368 five blanks, five options, and a relevant animation of the term/concept. Students drag
369 each of the five phrases to the correct blanks within the definition. If the dragged
370 phrase is not correct, it snaps back to its original place. When the blanks are correctly
371 filled, a congratulation message pops up and displays the complete definition of the
372 term.

373 • *Formulas*. In collaboration with the physics experts, we created annotated mathe-
374   matical formulas for the physics terms. Clicking on the formula option reveals the
375   formula, along with a short explanation of each component/variable.
376 • *Hewitt videos.* Hewitt videos allowed students to watch a short (1–2 min) physics
377   video developed by Paul Hewitt explaining the primary concept related to the
378   level. The physics experts helped select the most relevant videos for the game
379   competencies. With Paul Hewitt's permission, the team edited the length of the
380   videos to make them illustrate a targeted competency.
381 • *Glossary*. The glossary provides brief explanations of 28 physics terms. The terms
382   have been selected, edited, and revised by the physics experts.

383   **Show me a Hint or Solution**. Clicking on this option takes the student to either
384   a worked example or a hint—both of which are linked to the specific level being
385   played.

386   • *Worked examples.* Worked examples are videos of expert solutions of game levels.
387   All worked examples are less than a minute long with the majority being less than
388   30 s. We created at least one worked example for each game level and solution
389   agent (130 + levels—both task types). From our first and second usability studies,
390   we found that students liked worked examples and selected this support more
391   frequently than any of the other types. However, this support enabled students
392   to solve levels without thinking or problem solving first. Consequently, our new
393   incentive system (discussed in detail later) charges for viewing this support.
394   • *Hints.* In the first version of PPv2, this support was called *Advice.* When this support
395   was selected, it triggered a short, general hint for solving a level (e.g., "Remember
396   that a larger force will cause an object to accelerate faster"). Results of the first
397   usability test showed this support was not effective. Students commented that the
398   advice was too vague and thus unhelpful. So, we replaced the original advice with
399   level-specific physics solution hints (e.g., "Try drawing a ramp").

400   **Show me Game Tips**. If students are playing the game for an extended period of
401   time, they will likely forget some of the game mechanics and ways to draw different
402   simple machines (e.g., ramp or lever). Consequently, we developed a support related
403   to gameplay—show me game tips. When students select this support, a window
404   opens with tabs that each contains game play reminders (Fig. 9).

405   • *"Controls" and "Simple Machines."* These only appear when the player is in a
406   sketching level. When a student clicks on the "Controls" tab, a scrollable page pops
407   up showing game mechanics (i.e., nudge, draw an object, and delete an object for a
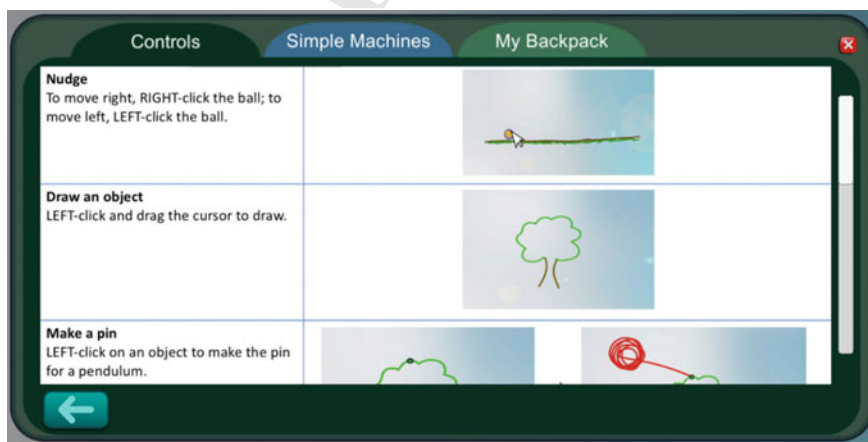408   sketching level, etc.). When a student clicks on the "Simple Machines" tab, images

Fig. 9 "*Show me Game Tips*" menu in *PPv2*

of the four simple machine tutorials (i.e., lever, pendulum, ramp, and springboard) appear. Each image is clickable and can be enlarged. By viewing the larger images, learners can quickly see how to create the agents without going through the full tutorials again.

- *Tools*. This option only appears when the player is in a manipulation level. Here players view rules for the sliders and a short explanation about other tools available (i.e., puffers and blowers).
- *My Backpack*. In both sketching and manipulation levels, "Show me Game Tips" includes "My Backpack." A screenshot from "My Backpack" will be shown with textboxes pointing at different parts of "My Backpack" explaining the various functions.

Incentive System. To encourage student performance and use of learning supports, we added an incentive system in *PPv2*. Most of the incentive system is contained within *My Backpack* (accessed via the top left corner of the level selection area in *PPv2*). When clicked, *My Backpack* provides information about progress in the game, as well as a space to customize game play (Fig. 10). That is, two progress bars—one for sketching levels and one for manipulation levels—show how many levels the student has solved and how many remain. A money bag displays their current balance with a drop-down function that shows the amount of gold and silver coins they have collected so far. The "Physics" tab shows the estimated competency level for each targeted physics concept (based on real-time stealth assessment), and the Store tab provides options to change the background music, background image, or ball type. This customization is an additional component of the incentive system and must be purchased by students with the money they make in the game.

Each level in the game has a "par" that is based on the degree of difficulty of the level. Each level was scored on two difficulty indices, game mechanics and physics concepts. A composite score was used to create the par. For sketching levels, the par is based on the minimum number of objects used in a solution. For manipulation levels, the par is based on attempts (i.e., each time a slider adjustment is made and the "Play" button clicked). If the player's solution is at or under par, a gold coin (worth $20) is given, and otherwise, a silver coin (worth $10) is awarded to the player. In Fig. 11, you can see that the player has collected eight gold coins and two silver coins, and the amount of money is $180.

AQ2

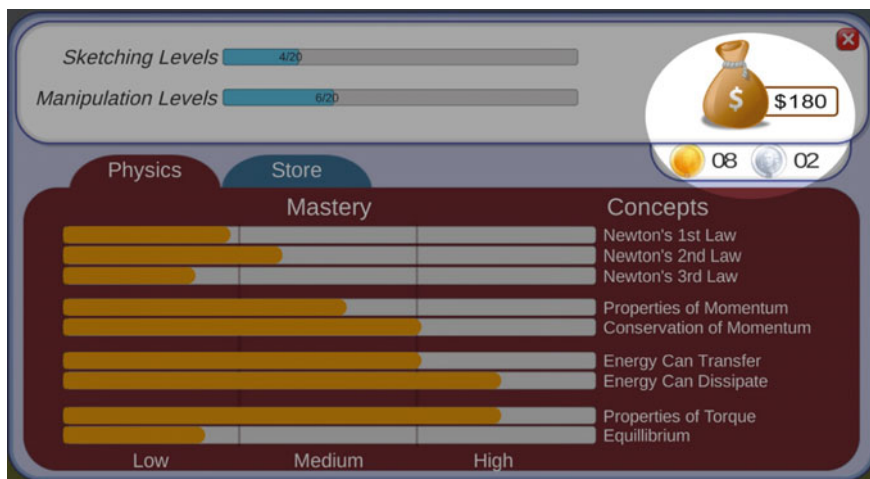**Fig. 10** My Backpack views—physics estimates (left) and store (right)

Fig. 11 Money bag and coins collected in *PPv2*

4 Testing the New Supports and Test Items—a Usability Study

The purpose of our most recent usability study was to investigate the effectiveness of the new animations when combined with gameplay, and pilot test a set of new near-transfer test items we developed as an external measure of physics understanding. For these purposes, we selected two minimally overlapping concepts in our competency model: energy can transfer (EcT) and properties of torque (PoT).

4.1 New Learning Supports—Physics

The new learning supports we included in *PPv2* for this study consist of seven new physics animations explaining the EcT and PoT concepts. The production of these supports was an outcome of our previous usability studies.

4.2 Measures

Physics Understanding Test. We created two physics test forms (Form $A = 14$ items; Form $B = 14$ items) each of which included 10 near-transfer test items (new for this study), and 4 far-transfer test items (used in prior studies). Each item included in the test targeted either EcT or PoT (see Figs. 12 and 13 for examples of a near- and far-transfer item).
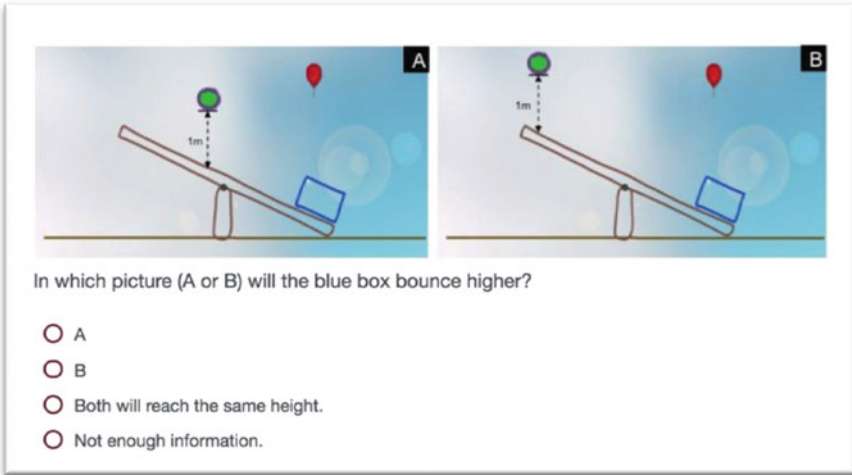
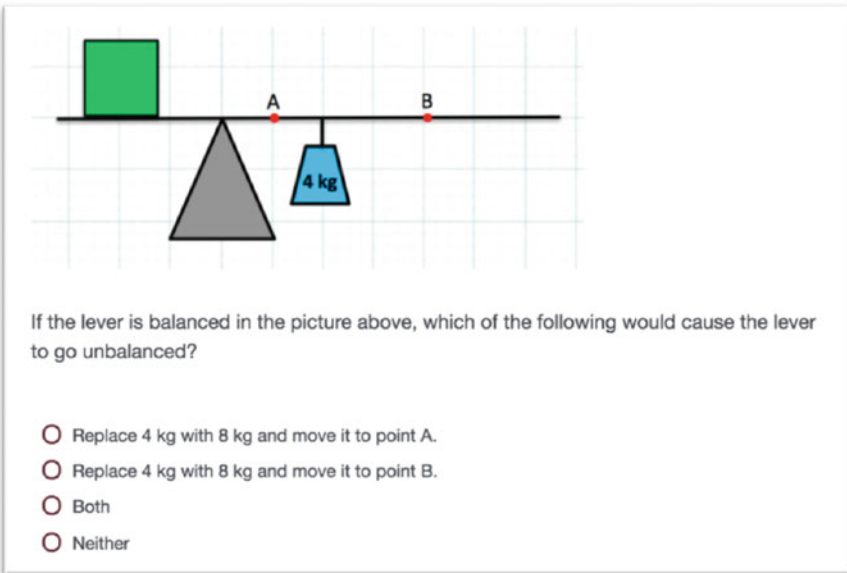**Fig. 12** An example of our PoT near-transfer test items. The answer is B

Fig. 13 An example of our PoT far-transfer test items. The answer is B

459 **Game and Learning Supports Satisfaction Survey**. To evaluate students' sat-
460 isfaction of the game and our new learning supports, we used a 16-item Likert-scale
461 questionnaire, developed in house, with two parts: (1) game satisfaction and (2)
462 learning supports' satisfaction.

4.3 Method

464 **Participants**. Our convenience sample included 14 students (6 seventh graders, 8
465 eighth graders; 6 females, and 8 males) from the School of Arts and Sciences (SAS)
466 in Florida. They were compensated with a $10 gift card upon completion of the study.
467 All students played the same version of the game.

468 *PP* **Levels Selected**. We selected 30 sketching levels (a mixture of levels with
469 PoT or EcT as their primary physics concept) with variable difficulty levels. We also
470 included the new set of sketching tutorial levels. In total, students had 35 levels to
471 complete.

472 **Procedure**. Students first completed an online demographic questionnaire fol-
473 lowed by the pretest (Form *A*). Next, students played the game, individually, for
474 75 min. Student gameplay was monitored by six researchers. The researchers allowed
475 students to access the learning supports (worked examples, physics animations, and
476 game tools) freely during the first 20 min. For the following 55 min, students were
477 only allowed to access the "physics supports" (i.e., our new animations), and the
478 researchers prompted the students to access them every 8 min or after completing
479 three game levels. At the end of the 75 min of gameplay, students completed the
480 posttest (Form *B*) and the game and learning supports satisfaction questionnaire.

4.4 Results

482 Despite the limitations of this usability study (i.e., small sample size, short gameplay
483 time, and lack of control group), we obtained some useful findings that can help us
484 improve the game for future, larger studies. We first examined the near-transfer items
485 and identified a few problematic items. Then, we examined the mean differences
486 between the various subsets of the pretest and posttest. Finally, we looked at the
487 game and learning supports satisfaction questionnaire to see how the students felt
488 about the game in general and the learning supports in particular.

489 **Item Analysis**. Cronbach's α for the EcT near-transfer items (both pre- and
490 posttest items) was 0.61, and the α calculated for the PoT near-transfer items (pre-
491 and posttest items) was 0.38. We found three items with zero mean variability (either
492 all the students got those items wrong or right) and three items showing near-zero
493 mean variability (only 1 or 2 students got those items right). These items have been
494 revised for future use. It is expected that when we pilot test these revised items and
495 have a larger sample size, we will obtain a higher reliability for these items.

Physics Understanding. To assess students' physics understanding, we analyzed the pretest and posttest relative to their sections as follows: (1) near-transfer EcT tests scores, (2) near-transfer PoT test scores, (3) overall near-transfer test scores (with both EcT and PoT items combined), (4) overall far-transfer test scores, and (5) overall pretest and posttest scores with all the items included (near and far-transfer). Then we conducted several paired-sample t-tests to examine the differences between the means coming from these subsets, and several correlational analyses to examine the relationships between these subsets in the pretest and posttest. Table 1 summarizes our findings.

As shown in Table 1, students scored significantly higher on the posttest compared to the pretest ($M_{pre} = 0.57$, $M_{post} = 0.63$, $t(13) = -2.20$, $p < 0.05$, Cohen's $d = 0.60$). In addition, the near-transfer pretest significantly correlated with the near-transfer posttest ($r = 0.53$, $p < 0.05$).

Game and Learning Supports Satisfaction. To get a sense about students' overall satisfaction from the game and the learning supports, we analyzed responses to the questionnaire which followed the posttest. We divided the results into two parts: game satisfaction (Likert-scale, 1 = strongly disagree to 5 = strongly agree; see Table 2), and learning supports satisfaction (Likert-scale, 1 = strongly disagree to 5 = strongly agree; see Table 3).

As shown in Table 2, students really liked the game on average ($M = 4.24$, $SD = 0.62$). This finding is consistent with our previous findings in other research studies (e.g., [40]). Also, students agreed that the game helped them learn some physics ($M = 3.93$, $SD = 1.07$).

Table 3 shows that students found the learning supports satisfying and useful ($M = 3.99$, $SD = 0.51$) and reported the new animations helped them learn physics ($M = 3.79$, $SD = 1.19$). Moreover, males and females equally enjoyed the game and the supports.

Having a small sample size and one-group pretest–posttest design can only provide preliminary insights. The overall results from this usability study suggest we are

Table 1 Descriptive statistics, paired-sample *t*-tests, and correlations for physics measures ($n = 14$)

	Pretest		Posttest		Paired-sample *t*-test (pre and post)		Correlation (pre and post)	
Measures	*M*	*SD*	*M*	*SD*	*t (13)*	sig.	*r*	sig.
EcT	0.44	0.25	0.54	0.16	−1.71	0.11	0.51	0.06
PoT	0.76	0.16	0.76	0.22	0.00	1.00	0.20	0.49
Near transfer	0.60	0.12	0.65	0.18	−1.61	0.13	0.53	0.04*
Far transfer	0.48	0.15	0.57	0.18	−1.44	0.17	0.05	0.87
All items	0.57	0.07	0.63	0.09	−2.20	0.04*	0.22	0.44

Note The means are standardized averages
*Significant at the $p < .05$. EcT = near-transfer EcT items. PoT = near-transfer PoT items

Table 2 Likert-scale game satisfaction questionnaire ($n = 14$)

Items	M	SD
I enjoyed the game very much	4.57	0.85
I thought the game was boring (RC)	4.71	0.83
The game did not hold my attention (RC)	4.29	1.20
I thought I performed well in the game	4.00	0.56
I was pretty skilled at playing the game	3.71	0.83
I put a lot of effort into solving levels	4.43	0.76
The game helped me learn some physics	3.93	1.07
Physics is fun and interesting	4.36	1.15
I'd like to play this game again	4.21	1.19
I'd recommend this game to my friends	4.14	1.29
Game satisfaction scale	4.24	0.62

Note RC = reverse coded

Table 3 LS satisfaction questionnaire ($n = 14$)

Items	M	SD
The "level solutions" helped me solve the levels	4.14	0.86
The "physics supports" helped me learn physics	3.79	1.19
The supports were generally annoying (RC)	4.14	1.23
The supports were pretty easy to use	4.21	0.70
The supports did not help me at all (RC)	4.00	1.18
I'd rather solve levels without supports (RC)	3.64	1.50
LS satisfaction scale	3.99	0.51

Note RC = reverse coded

on the right path. However, we have revised our near-transfer items (based on item analysis results) and will conduct more pilot testing on those items before using them in larger studies. Also, we will collect more qualitative data on our new learning supports with further rounds of revisions as needed. The reflection on students' learning experiences prepares us for the next phase of the project—implementing an adaptive algorithm into the game. Next, we discuss the remaining steps needed to include adaptation using game-based learning analytics in *PPv2*.

5 Testing Game-Based Learning Analytics in *Physics Playground*

Shute, Ke, and Wang [17] listed ten steps—derived from multiples studies conducted relative to stealth assessment—to include accurate measurement and adaptation in PP:

1. Develop the full competency model (CM) of the targeted knowledge, skills, or other attributes based on full literature and expert reviews
2. Select or develop the game in which the stealth assessment will be embedded
3. Identify a full list of relevant gameplay actions/indicators/observables that serve as evidence to inform CM and its facets
4. Design and develop new tasks in the game, if necessary
5. Create a *Q*-matrix to link actions/indicators to relevant facets of target competencies to ensure adequate coverage (i.e., enough tasks per facet in the CM)
6. Establish the scoring rules to score indicators using classification into discrete categories (e.g., solved/unsolved, very good/good/ok/poor relative to quality of the actions). This becomes the "scoring rules" part of the evidence model (EM)
7. Establish statistical relationships between each indicator and associated levels of CM variables (EM)
8. Pilot test Bayesian networks (BNs) and modify parameters
9. Validate the stealth assessment with external measures
10. Include adaptation of levels and/or support delivery in the game.

At the time of writing this chapter, we have completed steps 1 through 8 with the new version of PP. That is, we have revised/elaborated the competency model of physics understanding, (b) created task types and associated levels that provide the evidence we need to assess students' physics understanding via stealth assessment, (c) developed and tested a variety of learning supports to help students enhance their physics knowledge during gameplay, and (d) set up an incentive system that can boost students' motivation to use the learning supports in the game. In the coming months, to complete the 10-step guideline mentioned above, we will add and test online adaptation [41] in *PP* for the selection of levels and learning supports delivery.

Level Selection. During gameplay, students provide a plethora of data (stored in a log file). The data are analyzed by the evidence identification (EI) process—in real time. The results of this analysis (e.g., scores and tallies) are then passed to the evidence accumulation (EA) process, which statistically updates the claims about relevant competencies in the student model—e.g., the student is at a medium level regarding understanding the concept of Newton's first law of motion. Using the stealth assessment results in *PP*, and based on an adaptive algorithm (see [19]), the system will pick the next level for the student. The best next level for a student is one with a fifty-fifty chance of success based on the student's prior performance in the game. In other words, the next level presented to the student will likely be in his/her ZPD [29].

573   **Learning Supports Delivery**. Currently, and in line with the game design notion
574   of learner autonomy in game play, we allow players to access the help voluntarily.
575   We will be testing the incentive system in an upcoming study, to see if it works as
576   intended (i.e., fosters use of physics supports and reduces abuse of worked exam-
577   ples). However, we have also developed a quit-prediction model that uses gameplay
578   data in the log file as the basis to make inferences about when a player is seriously
579   struggling and about to quit the game [42]. The model is based on high-level intuitive
580   features that are generalizable across levels, so it can now be used in future work
581   to automatically trigger cognitive and affective supports to motivate students to pur-
582   sue a game level until completion. To move toward game-directed learning support
583   adaptivity, we plan to include some simple rules that accompany the quit-prediction
584   model to determine when to deploy supports and which supports to choose.

6   Conclusion

586   Designing learning games, capable of assessing and improving student learning,
587   has serious challenges. For one, integrating just-in-time learning supports that do
588   not disrupt the fun of the game is a hurdle we are actively trying to surmount. In
589   this chapter, we discussed the importance of including learning supports and their
590   influence on learning and performance in educational games, described our own
591   experiences with stealth assessment and the design and development of different
592   learning supports in *PP*, presented the results from a recent usability study examining
593   the effects of our new supports on learning (with promising results on our new
594   learning supports and game satisfaction), and provided insights into the next steps
595   of game-based learning analytics via stealth assessment. Finally, we will continue
596   to design, develop, and test adaptivity of game levels students play in *PP* and the
597   learning supports they receive.

598   The central research study in our design and evaluation of learning support com-
599   ponents, including adaptive sequencing, is expected to yield principles that designers
600   of other educational games can use. Again, we aim to come up with a methodology
601   for developing game-based assessments and dynamically linking those assessments
602   to in-game learning supports. As we formalize the design process and share it, other
603   researchers and designers are able to utilize the methodology.

604   Through the use of game-based learning and stealth assessment, learning analytics
605   can be used to both measure *and* support student learning in an engaging way.
606   Harnessing the power of data generated by students in game play activities enables
607   more accurate assessments of student understanding and misconceptions than one-
608   off summative evaluations (e.g., final score). Better estimations of student struggles
609   and achievements can lead to better individualized instruction and more motivated
610   students, paving the way for new educational paradigms that replace the school-as-
611   factory metaphor.

615 References

616 1. Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper
617 and Row.
618 2. Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in
619 video games*. Cambridge, Massachusetts: The MIT Press.
620 3. Ravitch, D. (2000). *Left back: A century of failed school reforms*. New York: Simon & Schuster.
621 4. Shute, V.J. (2007). Tensions, trends, tools, and technologies: Time for an educational sea
622 change. In: C. A. Dwyer. (Eds.), The future of assessment: Shaping teaching and learning.
623 (pp. 139–187). Lawrence Erlbaum Associates, Taylor & Francis Group, New York, NY.
624 5. Kliebard, H. (1987). *The struggle for the American curriculum, 1893–1958*. New York: Rout-
625 ledge and Kegan Paul.
626 6. Bobbitt, J. F. (1912). The elimination of waste in education. *The Elementary School Teacher.,
627 12,* 259–271.
628 7. Lemann, N. (1999). The IQ meritocracy: Our test-obsessed society has Binet and Terman to
629 thank—or to blame [Electronic version]. *Times, 153,* 115–116.
630 8. Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and
631 learning: A Systematic review and meta-analysis. *Review of Educational Research., 86,* 79–122.
632 9. Castell, S. D., Jenson, J. (2003). Serious play: Curriculum for a post-talk era. *Journal of the
633 Canadian Association for Curriculum Studies, 1.*
634 10. Prensky, M. (2001). *Digital game-based learning*. New York: McGraw-Hill.
635 11. Habgood, M. P. J., Ainsworth, S. E., & Benford, S. (2005). Endogenous fantasy and learning
636 in digital games. *Simulation & Gaming., 36,* 483–498.
637 12. Ritterfeld, U., Shen, C., Wang, H., Nocera, L., & Wong, W. L. (2009). Multimodality and inter-
638 activity: Connecting properties of serious games with educational outcomes. *CyberPsychology
639 & Behavior, 12,* 691–697.
640 13. Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students'
641 autonomous motivation on learning organic chemistry: A self-determination theory perspective.
642 *Science Education, 84,* 740–756.
643 14. Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional
644 support in game-based learning. *Computers & Education, 60,* 412–425.
645 15. Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S.
646 Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, NC:
647 Information Age Publishers.
648 16. Mislevy, R. J., Steinberg, L. S., Almond, R. G.(2003). Focus article: On the structure of edu-
649 cational assessments. *Measurement: Interdisciplinary Research & Perspective, 1,* 3–62.
650 17. Shute, V., Ke, F., & Wang, L. (2017). Assessment and adaptation in games. In P. Wouters & H.
651 van Oostendorp (Eds.), *Instructional techniques to facilitate learning and motivation of serious
652 games* (pp. 59–78). Cham: Springer.
653 18. DiCerbo, K. E., & Behrens, J. T. (2012). Implications of the digital ocean on current and future
654 assessment. In R. Lissitz & H. J. Jiao (Eds.), *Computers and their impact on state assessment:
655 Recent history and predictions for the future* (pp. 273–306). Charlotte, NC: Information Age.
656 19. Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing it—Or
657 can you? Evaluating an assessment for learning system called ACED. *International Journal of
658 Artificial Intelligence in Education, 18,* 289–316.

20. Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., D'Mello, S. Toward a generalized competency model of collaborative problem solving. Computers & Education. (in press).

21. Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education, 70,* 65–79.

22. Lang, J., O'Neil, H. (2008). *The effect of presenting just-in-time worked examples for problem solving in a computer game.* In Presented at the American Educational Research Association Conference, New York.

23. Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 13,* 423–451.

24. Leutner, D. (1993). Guided discovery learning with computer-based simulation games: Effects of adaptive and non-adaptive instructional support. *Learning and Instruction, 3,* 113–132.

25. O'Rourke, E., Ballweber, C., Popovií, Z.(2014). Hint systems may negatively impact performance in educational games. In *Presented at the Proceedings of the first ACM conference on Learning @ scale conference*. Atlanta, GA, USA March 4.

26. Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15,* 313–331.

27. Moreno, R., & Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology, 94,* 598–610.

28. Cowley, B., Charles, D., & Black, M. (2008). Toward an understanding of flow in video games. *Computers in Entertainment, 6,* 1–27.

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

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

31. Kickmeier-Rust, M. D., & Albert, D. (2010). Micro-adaptivity: Protecting immersion in didactically adaptive digital educational games: Micro-adaptivity in digital educational games. *Journal of Computer Assisted learning, 26,* 95–105.

32. Shute, V. J., Graf, E. A., Hansen, E. G.(2005). Designing adaptive, diagnostic math assessments for individuals with and without visual disabilities. In L. PytlikZillig, R. Bruning, M. Bodvarsson (Eds.), *Technology-based education: Bringing researchers and practitioners together*. pp. 169–202. Information Age Publishing, Greenwich, CT.

33. Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research., 78,* 153–189.

34. Peirce, N., Conlan, O., Wade, V.(2008). Adaptive educational games: Providing non-invasive personalised learning experiences. In: *2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning*, pp. 28–35. IEEE, Banff, AB, Canada.

35. Hatano, G. (1982). Cognitive consequences of practice in culture specific procedural skills. *Quarterly Newsletter of the Laboratory of Comparative Human Cognition., 4,* 15–18.

36. Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education, 24,* 335.

37. Shute, V., Ventura, M., Kim, Y. J., & Wang, L. (2014). Video games and learning. In W. G. Tierney, Z. Corwin, T. Fullerton, & G. Ragusa (Eds.), *Postsecondary play: The role of games and social media in higher education* (pp. 217–235). Baltimore, MD: John Hopkins University Press.

38. Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., et al. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education, 86,* 224–235.

39. Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Bulletin, 30,* 415–433.

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

41. Lopes, R., & Bidarra, R. (2011). Adaptivity challenges in games and simulations: A survey. *IEEE Transactions on Computational Intelligence and AI in Games., 3,* 85–99.
42. Karumbaiah, S., Baker, R. S., Shute, V. (2018). *Predicting quitting in students playing a learning game*. In 11th International Conference on Educational Data Mining, pp. 1–10. Buffalo, NY.