

# Experience and Explanation: Using Videogames to Prepare Students for Formal Instruction in Statistics

Dylan A. Arena · Daniel L. Schwartz

Published online: 5 December 2013  
© Springer Science+Business Media New York 2013

**Abstract** Well-designed digital games can deliver powerful experiences that are difficult to provide through traditional instruction, while traditional instruction can deliver formal explanations that are not a natural fit for gameplay. Combined, they can accomplish more than either can alone. An experiment tested this claim using the topic of statistics, where people’s everyday experiences often conflict with normative statistical theories and a videogame might provide an alternate set of experiences for students to draw upon. The research used a game called *Stats Invaders!*, a variant of the classic videogame *Space Invaders*. In *Stats Invaders!*, the locations of descending alien invaders follow probability distributions, and players need to infer the shape of the distributions to play well. The experiment tested whether the game developed participants’ intuitions about the structure of random events and thereby prepared them for future learning from a subsequent written passage on probability distributions. Community-college students who played the game and then read the passage learned more than participants who only read the passage.

**Keywords** Computer games · Statistics instruction · Assessment · Interactive learning environments

## Introduction

### Experience and Explanation

When people speak of “learning,” they often use the term in an undifferentiated way, much as they might use the term “food” rather than proteins, fats, or carbohydrates. Petrich et al. (2013) note that when educators and policymakers observe children who are enthralled by a videogame or museum exhibit, they often ask, “But are they learning?” (p. 50). A more meaningful question would be, “What are they learning?” This rephrasing acknowledges that there are many different types of learning experiences and outcomes: Observation leads to imitation (e.g., Bandura et al. 1961), for example, and reinforcement leads to repetition (e.g., Skinner 1986). What the educators and policymakers probably mean to ask is, “What are the children learning that is valued in school?”

When considering the value of informal learning for school, it is a mistake to expect that informal experiences will produce the same learning outcomes as school—unless, of course, one makes the informal experience more school-like (e.g., Anderson et al. 2000). School emphasizes explanation: Students receive declarative accounts of facts and procedures. In contrast, informal education emphasizes experience: Videogames, museums, and science camps are designed to provide compelling experiences. Experiential and explanatory learning are different. If one uses a school-based test to evaluate what children have learned from a videogame, the results are likely to be disappointing (NRC 2011), because school tests emphasize explanatory knowledge.

However, informal learning may still have a valuable role to play for explanatory outcomes. Expository passages and lectures—the stuff of school—are good for delivering

---

D. A. Arena (✉) · D. L. Schwartz  
Stanford University, 485 Lasuen Mall, Stanford, CA 94305,  
USA  
e-mail: darena@alumni.stanford.edu

D. L. Schwartz  
e-mail: danls@stanford.edu

explanations. By the same token, these delivery mechanisms are often too compact to provide sufficient experiences for students to make full sense of the words and symbols that the explanations provide. Informal learning activities can help, because they excel at delivering experiences. For instance, the videogame *Civilization* transports players to a world where they make many complex choices while participating in a rich narrative. Videogames like this are unlikely to be sufficient for learning normative explanatory theories, but by providing relevant experiences, they may help prepare people to learn those theories. In other words, videogames can provide the experiences that expository teaching counts on, and expository teaching can provide the explanations that are hard to deliver in videogames. Rather than expecting a videogame to be a stand alone learning solution, one can conceptualize it within a larger learning ecology that links informal and formal learning experiences (e.g., Barron et al. 2012; Ito et al. 2012).

The goal of the current paper is to show that the experiences of an arcade-style videogame can prepare students to learn explanations provided through formal instruction. We make our demonstration in the context of college students learning about statistical distributions. This is a topic where students experience the world in non-normative ways and where a videogame might provide better experiences. However, the importance of our demonstration is not about teaching statistics per se. Our empirical goal is to reveal the hidden potential of videogames for academic learning outcomes by (a) demonstrating a technique for evaluating whether a videogame is providing effective experiential learning and (b) showing that videogames do not need to include expository content to be effective for academic outcomes that go beyond rote memorization.

### Preparation for Future Learning

Because videogames do not normally deliver normative explanations, evaluating them directly with an explanatory test would be a mismeasurement. Most assessments of learning use what Bransford and Schwartz (1999) called sequestered problem solving (SPS). In sequestered problem solving, students are shielded from contaminating sources of information that might enable them to learn during the test. An SPS assessment is appropriate for full-blown declarative and procedural knowledge, but it is a mismatch for experiential learning. In experiential learning, students develop intuitions that are likely to be tacit and inchoate. They are important experiences, but they may not have the verbal mediation that translates into answering abstract or general questions.

Bransford and Schwartz (1999) proposed an alternative measurement paradigm, which they called preparation for

future learning (PFL). In a PFL test, students receive learning resources as part of the overall assessment, and the question is whether their prior experiences prepare them to learn from these resources. For example, to evaluate whether a videogame experience prepares students to learn subsequent explanations, a PFL assessment would include expository material as part of the assessment and measure whether students learn from it.

Prior research has indicated that PFL assessments can capture the benefits of innovative models of instruction that SPS measures miss. For instance, Schwartz and Martin (2004) demonstrated that guided discovery experiences for learning about variance looked ineffective when compared to more traditional instruction using SPS measures. However, when evaluated by how well the experiences prepared students to learn a new statistical concept from a worked example, the discovery condition did twice as well as the traditional-instruction condition. Similarly, Schwartz and Bransford (1998) found that asking students to analyze raw data from psychological experiments did not appear useful compared to summarizing a relevant passage. Yet when all the students received a follow-up lecture, the students who had analyzed the data did much better on a posttest that required predicting the outcomes of a novel but related psychological experiment. In the PFL model of assessment, expository resources such as lectures, readings, and worked examples are part of the assessment from which students can learn. PFL is a dynamic assessment (Feuerstein 1979) rather than a summative one.

When digital games provide well-designed experiences, they may prepare students for future learning from formal explanations. In this scenario, games do not have the burden of teaching the formal content, which can be difficult to accomplish through game mechanics. Instead, the game can prepare students to learn the formal content later. In one study, for example, community-college students were randomly assigned to play one of two commercial videogames in their homes for 15 h over several weeks, while a control group did not play either game (Arena 2012). The two games, *Civilization IV* and *Call of Duty 2*, provide experiences that are relevant to World War II, but neither game was designed to teach anything in particular about the war. Nonetheless, students assigned to play these games learned more from a subsequent lecture about World War II than did the control group. A further indication of the influence of the specific game experiences is that those students who played *Civilization IV* were more likely to focus on nation-level issues discussed in the lecture, whereas students who played *Call of Duty 2* were more likely to focus on local tactics. In sum, the games provided students a rich body of prior knowledge that could help them grasp and elaborate the content of the lecture.

In the following study, we compare SPS and PFL measures for evaluating the learning outcomes of a

videogame that we built to help teach statistics. Two groups of students played one of two versions of the videogame, and one group of students did not play the game. Afterward, half of each group of students received an expository lesson about statistics and half did not. We then measured their overall learning gains with a posttest of explanatory knowledge. The prediction was that students who only played the game would do poorly, whereas students who played the game and received the passage would outperform students who only read the passage. This study replicates earlier studies demonstrating that PFL assessments can detect the benefits of experience for subsequent learning from explanations. The study also goes beyond the prior studies because it used an arcade-style game that is heavy on fast-action experience and very light on verbal mediation compared to the prior studies and tried to teach in a domain where people have pervasive misconceptions.

### The Challenge of Statistics

For many disciplines, people do not experience the world in ways that match modern theories. For instance, people do not experience evolution; they do not experience a rotating earth; they even do not experience the frictionless world imagined by physicists. Instead, they have thousands of hours of experience that can be misaligned with those theories. The surfeit of misaligned experiences can lead to learning challenges and persistent misconceptions.

Videogames may help, because they can orchestrate sustained experiences that are more closely aligned with the explanations of experts. In the current project, we developed a game that would prepare students to learn statistical concepts. The domain of statistics is notorious for persistent misconceptions that are difficult to remediate through traditional instruction (Nisbett et al. 1983). Tversky and Kahneman (1974) offer a nice summary of people's failures in this realm, along with the following conclusion:

Although everyone is exposed, in the normal course of life, to numerous examples from which these rules could have been induced, very few people discover the principles of sampling and regression on their own. Statistical principles are not learned from everyday experience because the relevant instances are not coded appropriately. For example, people do not discover that successive lines in a text differ more in average word length than do successive pages, because they simply do not attend to the average word length of individual lines or pages. Thus, people do not learn the relation between sample size and sampling variability, although the data for such learning are abundant (p. 1130).

The excerpt highlights that people do not interpret their experiences of probabilistic situations in ways that support the development of normative interpretations of those situations. This problem is exemplified by what Konold (1989) calls the *outcome-oriented approach*. People who adopt this approach believe that the task in probabilistic situations is to predict the outcome of a specific instance (rather than to characterize a distribution of instances), a belief that can easily lead people astray. For example, Konold presented educated adults with a normal six-sided die, five faces of which had been painted black. He then asked whether rolling the die six times would be more likely to yield six black results or five black results and one white result. Participants with an outcome approach tended to reason that, because a black result is the most likely result for each roll, the best prediction would be six black results (getting one white result and five black results is actually 30 times more likely than getting six black results). Much of people's everyday experience involves predicting single outcomes, usually through causal reasoning or by relying upon readily available memories. This, along with the lack of appropriate representations for encoding and thinking about collections of probabilistic events, makes it unsurprising that people tend to think in terms of single outcomes.

### A Game-Based Solution

There have been a number of excellent approaches to help people learn normative statistical concepts. These include well-designed simulations and visualizations (e.g., Konold 2007), materials that support productive discussion (e.g., Himmelberger and Schwartz 2007), and the delineation of optimal learning progressions (e.g., Lehrer et al. 2013). Here, we took a videogame approach. We do not propose that our solution is better than other solutions, and the current research is not an attempt to find out. Rather, the goal is to show that videogame experiences can improve student preparation for learning—in this case from a simple passage, but conceivably from many of the instructional materials that other people have developed.

A videogame has the potential to remediate people's confusions about statistical outcomes by offering sustained experiences that facilitate thinking about probability distributions and a game mechanic that requires them to do so. If the game is fun, people may play long enough to build normative intuitions about the characteristics of probability distributions, which could allow future instruction to resonate with those experiences rather than to conflict with the overabundance of experiences focused on predicting single outcomes.

In the domain of statistical reasoning, the histogram is the fundamental representation to help people think about

aggregates of chance events. The histogram and its continuous cousin, the probability density curve, can provide complete information about the behavior of a random variable. Interpreting histograms and probability density curves, however, does not make a great core mechanic for a game. Fortunately, a closely related concept in probability—repeated sampling from a population—maps into a well-tested game mechanic.

One of the first iconic videogames, *Space Invaders*, involved shooting alien spaceships descending from the sky. We borrowed this classic mechanic to make a new game called *Stats Invaders!*, which adds two new twists. First, the descending invaders follow a variety of probability distributions that determine where they fall from the sky. For example, when a normal distribution generates the alien attack, invaders are most likely to descend from the center of the screen, with less frequent descents from the edges. The second twist is that the player's task is not simply to shoot these invaders before they land but also to generalize from these individual observations to determine which of two probability distributions describes the invaders' overall pattern of attack. Our hope was that these two additions to the classic game would help students begin to understand intuitively that randomness does not mean without pattern, but rather that even random events have regular, identifiable patterns that can be expressed by distribution graphs.

### Game Design

*Stats Invaders!* gameplay is separated into levels embedded within stages (for technical details, see Arena and Schwartz 2010). Each level is a single opportunity to identify patterns of alien attack. Each stage is a collection of levels, and each new stage introduces increasingly difficult challenges. The five panels in Fig. 1 illustrate the progression of challenges in *Stats Invaders!* The game begins (Fig. 1a) with a “practice” stage, in which players learn to associate a pattern of alien attack with a distribution (a probability density curve, though they are not described as such). In the next stage (Fig. 1b), players must decide between two distributions that differ in shape (probability density function), center (mean), and spread (variance). The third stage (Fig. 1c) is like the second except that both distributions have the same center; and in the fourth stage (Fig. 1d), they have the same center and shape, differing only in spread (Figure 1d illustrates how many observations it can take to distinguish two distributions based only on a difference in variance: Each dot in the histogram below the player's ship represents one invader destroyed).

After the fourth stage (Fig. 1e), the game makes a conceptual shift from simple hypothesis testing (with the top distribution as  $H_0$  and the bottom distribution as  $H_1$ ) to

general hypothesis testing of a known  $H_0$  against a universe of possible alternatives—i.e., the classic paradigm of rejecting (or not) the null hypothesis, which is taught in every introductory statistics course. The fifth, sixth, and seventh stages are just like the second, third, and fourth, respectively, except that the bottom distribution is hidden, forcing players to decide whether the pattern of alien attack is different enough from the displayed distribution to “reject” that distribution in favor of the unknown alternative.

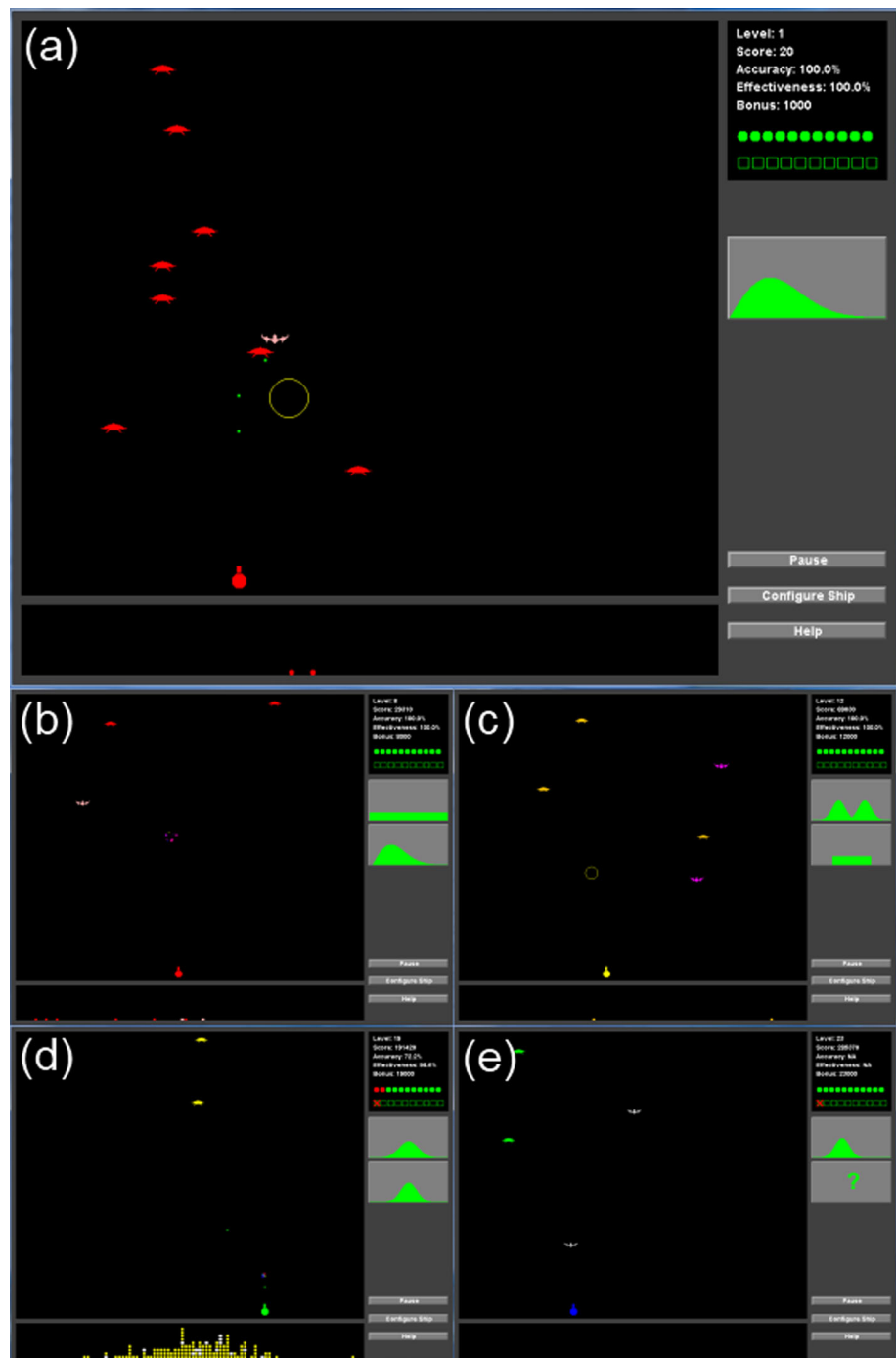
As expected in an arcade-style game, *Stats Invaders!* awards points to players for destroying alien invaders and for choosing the correct distribution on each level. Players earn points to unlock new ships that shoot and move more quickly. These ships become necessary to keep up with the pace of alien attack, which speeds up as the levels progress. Players are penalized for allowing invaders to reach the ground and for choosing the incorrect distribution. Enough of either of these mistakes will end the game.<sup>1</sup>

### Experiment

Our broader hypothesis is that the experiences of a well-designed videogame can prepare students to learn the explanatory content typical of schools. We tested this hypothesis by investigating whether our specific game prepared students to understand an expository explanation of statistical distributions. We asked community-college students to take a 10-item, open-response pretest about basic statistical-distribution concepts. We then randomly assigned them to either a no-game control condition or a

<sup>1</sup> One additional gameplay feature that is distinctive to *Stats Invaders!* is that players are eventually forced to choose between the two distributions displayed on the right of each screen. As each new invader is displayed on the screen, the game calculates the probability of having drawn this particular sequence of observations from each of the two displayed distributions. When the likelihood ratio reaches 500—i.e., when the sample of invaders is 500 times more likely to have been generated by one of the distributions than by the other—the game stops sending invaders and forces the player to make a choice. This game feature was added to combat what Edwards (1968) called *conservatism in information processing*, which is the tendency for people to require many more observations than are mathematically necessary to decide with high confidence between two alternative hypotheses (when players of *Stats Invaders!* are forced to make a choice, they often express the feeling that they are guessing, even though one distribution is 500 times more likely than the other to be the correct answer). We added this feature so that participants would not persevere on individual levels just to be absolutely sure they chose the correct distribution, but we also appreciate the feature for its ability to illustrate the phenomenon of conservatism when that phenomenon is presented to players of *Stats Invaders!* in subsequent formal instruction.

**Fig. 1** Stage progression from practice to general hypothesis testing. **a** Stage 1 is a practice stage to familiarize players with the distributions; **b** stage 2 has different shapes, centers, and spreads; **c** stage 3 has different shapes and spreads; **d** stage 4 has only different spreads; **e** stages 5, 6, and 7 repeat the progression of stages 2, 3, and 4, except with one distribution hidden



gameplay condition. (There were two versions of the game: distribution mode and proportion mode. In the “Methods” section, we explain the subtle difference between these two modes, but they are similar enough that we discuss them collectively here.) Crossed with the gameplay factor was a passage factor. The passage factor determined whether students read an explanatory passage about the statistical-distribution concepts covered in the pretest before taking a posttest, which was just a parallel

form of the pretest.<sup>2</sup> Students who read the passage completed a PFL assessment, whereas students who did not read the passage completed an SPS assessment because they did not have the learning resource. Thus,

<sup>2</sup> The passage did not cover all of the key experiences in the game, such as hypothesis testing or minimal sample size. The game is intended to support many weeks of statistics curriculum. Given the goal of an initial test of our main hypotheses and the software, we did not scale up to multiple weeks of instruction for this study.



students either (a) only played a game, (b) only read the passage, (c) did neither, or (d) did both. All students completed the 10-item pretest and the parallel 10-item posttest.

This experimental design addresses two overlapping questions. The first question is whether videogame experiences can prepare students to learn explanations characteristic of school. Our specific prediction is that students in the game conditions who then receive the passage will outperform students who only play the games, only read the passage, or do neither. The second question is whether a PFL assessment is better suited to evaluating the effects of videogame experiences than is an SPS assessment. We predict that students who play the game but do not receive the passage before the final test (SPS) will do worse than the students who play the game and receive the passage (PFL). To ensure that the PFL version of the assessment is detecting the value of the game, the game plus passage condition needs to outperform the passage-only condition, lest the learning can be attributable solely to having read the passage.

## Methods

### Participants

The participant pool comprised students enrolled in various introductory social-sciences courses at a community college in the San Francisco Bay Area during a single academic quarter. Of the 97 people who chose to participate in our study in exchange for course credit, we obtained usable data for 83 (35 females and 48 males).<sup>3</sup> Participant dropout was the result of a software bug that periodically crashed the computer script delivering the various components of the experiment (game, passage, and pre- and posttests). When participants' data were lost, we recruited more participants to take those spaces in the experimental design. By the end of our data-collection period, all conditions had  $n = 14$  (six females and eight males) except the no-game/passage condition, which had  $n = 13$  (five females). The median participant age was 20 years, but participants ranged from 17 to 52 years old, with five participants who were older than 30.

### Design

The experiment used a  $3 \times 2 \times 2$  factorial design. For the three levels of the gameplay factor, we randomly assigned participants to either play the game in one of two modes

(*distribution* mode or *proportion* mode) or not play at all. For the two levels of the passage factor, we randomly assigned participants to either read a passage about probability distributions or not read anything. We stratified the randomization procedure to balance gender across conditions. Finally, participants completed parallel pretest and posttest forms so we could compute individual gain scores.

### Materials

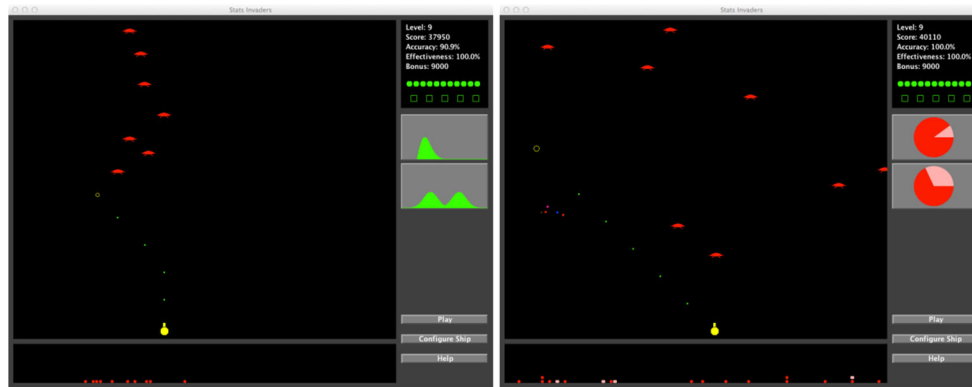
**Computer Program** The general description of the game was provided above. The game is written in Java, with simple graphics and gameplay reminiscent of early arcade videogames. Players control a spaceship to shoot a mixture of normal and “special” (faster and differently colored) alien invaders falling from the sky. Both the proportion of “special” invaders and the locations from which invaders fall are determined by probability distributions (generated by the stochastic simulation in Java framework, L’Ecuyer et al. 2002). The goal of each level is to decide which of two displayed patterns best reflects the alien attack.

The difference between the two game conditions involved the representation participants used to make choices about the pattern of alien attack. To pass a given level, participants had to choose between the two graphs on the right of the screen. The choice in distribution mode (Fig. 2, left) was between two possible (spatial) distributions, which required conceptualizing the shape of the alien attack as a probability density curve. In proportion mode (Fig. 2, right), the choice was between two proportions of “special” invaders, which required estimating the relative frequency of two types of invaders (and does not involve thinking in terms of distribution shapes). Gameplay in the two modes was otherwise identical. We included the two game conditions to determine whether including the formal representation of probability density curves within the gameplay would help build student intuitions about distributions as aggregate representations of outcome probabilities.

**Passage** Half of the participants read a two-page passage prior to the 10-item posttest. The passage explained that randomness has patterns, which makes it possible to predict collections of outcomes. It also included images of three distributions (uniform, normal, and skewed), along with brief descriptions about how to interpret the distributions and their implications. As an example, here is the paragraph that accompanied the skewed distribution:

The distribution in the middle of the picture is called **skewed**, because it’s sort of off-balance. This pattern is a good way to describe home prices in the Los Angeles area. The pattern has no values on the left,

<sup>3</sup> A power analysis conducted using the G\*Power software suite (Faul et al. 2009) suggested that a total sample size of 81 would be sufficient to detect a “medium” effect with 85 % power and  $\alpha = .05$ .



**Fig. 2** Gameplay in distribution mode (*left*) and proportion mode (*right*)

which tells us that there are no houses below a certain price. The pattern also has a very long “tail” to the right, which tells us that there are a few houses that are really, really expensive. (“Tails” of distributions tell us about the likelihood of observations that are far from what usually happens.) The way the pattern rises up in the middle tells us that the majority of houses have prices in this range of values, not too cheap or too expensive relative to other houses. In general, skewed distributions are good for describing random events where most values are near each other but a few are far away in only one direction (like the number of parking tickets a person has gotten: most people have gotten a few, but some people have gotten lots).

**Tests** We created two parallel forms of a 10-item free-response test covering the concepts presented in the passage about probability distributions. Items tested basic vocabulary (e.g., “What does it mean for an event to be ‘random’?”), understanding of graphical representations (e.g., “Given the distribution shown here, how likely is outcome A? Please explain your answer,” along with a labeled graph), and abilities to think in terms of patterns rather than single events (e.g., “Suppose you wanted to find out how good a particular weather forecaster’s predictions were. You observed what happened on 10 days for which a 70 % chance of rain had been reported. On 3 of those 10 days, there was no rain. What would you conclude about the accuracy of this forecaster?” This question was adopted from Konold 1989). We administered these tests via the same computer program that displayed the game. Participants had between 45 and 90 s to type responses to each item. Each participant received one form of the test as a pretest and the other form as a posttest, counter-balanced across treatments. All items were dichotomously scored, resulting in score ranges from 0 to 10 on each test.

## Procedure

Each experimental session had between two and ten participants sitting at standard classroom laptop computers. We placed the laptops roughly four feet apart on tables along two walls of a single room, with participants seated facing the walls. Once all participants had arrived, the experimenter obtained informed consent and gave each participant a code to enter into the computer program. That code determined both the participant’s experimental condition and which parallel forms of the test would serve as the pretest and posttest for that participant.

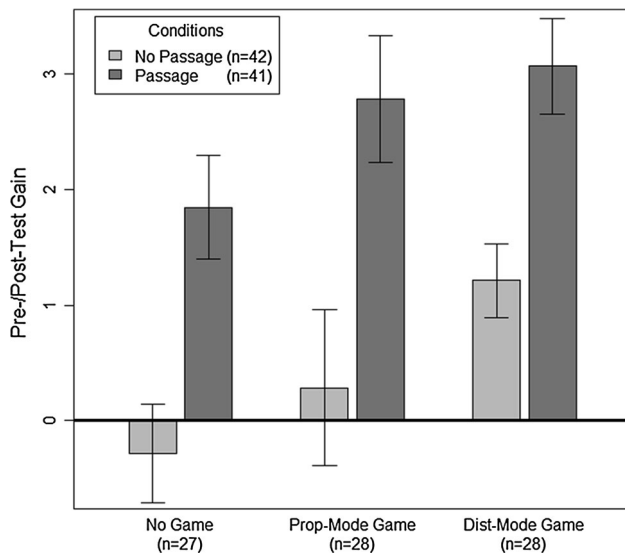
After all participants had received codes, the experimenter instructed them to begin. The computer script controlled each participant’s experimental flow as follows:

- Administer one form of the test as a pretest, allowing 11 min for the test.
- If the participant is in a gameplay condition, administer the game (in distribution or proportion mode, as appropriate), allowing 27 min for gameplay.
- If the participant is in a passage condition, present the passage, allowing 7 min to read the passage.
- Administer the other form of the test as a posttest, allowing 11 min for the test.

Because people could finish at different rates, the computer program instructed participants to remain at their stations and quietly browse the Internet after finishing. This way, other participants would not feel rushed to finish. The experimenter thanked and released participants as soon as everyone in the session finished.

## Results and Discussion

The two forms of the 10-point test were parallel. Collapsing across conditions, the mean difference between form A and form B was one one-hundredth of a point at pretest and five one-hundredths at posttest. The aggregated



**Fig. 3** Pre-/posttest gains by condition. All error bars indicate one SE of that condition's mean

test means were 3.4 out of 10 at pretest and 4.9 out of 10 at posttest. The lack of a ceiling effect and similarity of forms allow us to simplify the analysis by using pre- to posttest gain scores as the outcome of interest (the reliability estimates of the tests were somewhat low: Cronbach's  $\alpha$  was .60 for the pretest and .64 for the posttest, which should be expected for a test that exceeds student knowledge and therefore produces a lot of guessing).

Figure 3 shows the gain scores (posttest – pretest) broken out by condition. The effect sizes (Cohen's  $d$ ) for each treatment compared to the no-game/no-passage control condition are as follows: proportion-mode game and no-passage = .28; distribution-mode game and no-passage = 1.11; no-game and passage = 1.38; proportion-mode game and passage = 1.74; distribution-mode game and passage = 2.22.

Descriptively, playing the game (in either mode) and then receiving the passage led to higher gain scores than only reading the passage or only playing the game. This supports the idea that the game provided experiences that helped students learn the explanations in the passage. Moreover, the results demonstrated the value of the PFL assessment approach: Had we only measured the value of the videogame without including the passage, it would have appeared that the game experience was a poor use of time compared to simply reading the passage. This conclusion would have been most erroneous for the proportion-mode version of the game. Participants who only played the proportion-mode version of the game did not show much change from pre- to posttest and did very poorly compared to those students who only read the passage. In contrast, proportion-mode participants who had

an opportunity to read the passage descriptively outperformed students who only received the passage.

To test these effects, we began with a two-way ANOVA that crossed the three gameplay and two passage factors on gain scores.<sup>4</sup> There were significant main effects of both the gameplay factor,  $F(2, 79) = 4.00, p < .023$ , and the passage factor,  $F(1, 79) = 30.37, p < .0001$ , with no significant interactions. This double main effect indicates that combining the experience of the game with the explanation of the passage led to the best learning overall.

We also tested a number of specific, directional (one-tailed) hypotheses with a priori orthogonal contrasts. These hypotheses examined the major questions driving the work: two main questions and one subsidiary question. Our first main question was whether gameplay plus the passage would lead to better learning than only gameplay or only the passage. The ANOVA demonstrates this effect broadly, but it is imprecise regarding the two different versions of the game. We had hypothesized that because the distribution mode of the game accentuates the visual representation used to organize probabilities (histograms/probability density curves), the distribution-mode game would provide the best chance for demonstrating that gameplay plus the passage was better than either alone. As predicted, the relevant contrast showed that the distribution-mode game and passage treatment led to greater learning than did both the no-game and passage treatment,  $t(25) = 2.01, p = .028$ , and the distribution-mode game and no-passage treatment,  $t(26) = 3.57, p < .001$ . These results simply repeat the finding of the ANOVA, but specifically regarding the distribution-mode version of the game.

Our subsidiary question addresses our above-mentioned hypothesis that the distribution-mode version of the game was better than the proportion-mode version. A corollary of this hypothesis is that the distribution version of the game would help students learn from the passage more than the proportion version would. This prediction was incorrect; participants in the distribution-mode game and passage treatment did not significantly outperform participants in the proportion-mode game and passage condition,  $t(26) = .042, p = .34$ .

Our second main question was whether the PFL assessment would be more sensitive to the value of gameplay than the SPS assessment. We had predicted that game-based learning would look relatively good by the PFL assessment (where students read the passage) but would look poor by an SPS assessment (where students did not read the passage). The preceding analyses demonstrated the first fork of this prediction: The PFL assessment was effective in showing the benefit of the games when coupled with a passage,

<sup>4</sup> We confirmed that assumptions for ANOVA were met using the car analysis package (Fox and Weisberg 2011).



because gameplay plus passage fared better than passage alone. The second fork of the prediction requires testing whether the SPS assessment was insensitive to the benefits of gameplay. Here, the hypothesis is that the value of the game will not reveal itself when students do not have a chance to read the passage: Playing the game without the passage should appear no better than doing nothing. For this test, we again separated the proportion-mode and distribution-mode games. As hypothesized, the proportion-mode game and no-passage treatment were not statistically different from the no-game and no-passage treatment,  $t(26) = 0.72$ ,  $p = .24$ . Against hypothesis, however, participants in the distribution-mode game and no-passage condition did learn more than did those in the no-game and no-passage condition,  $t(26) = 2.82$ ,  $p < .005$ . We consider this discrepancy between the two game conditions in the “[General Discussion](#).”

## General Discussion

Our leading claim is that it is possible to improve learning through the combination of experiential videogames and formal explanations. We tested this claim in the context of statistical distributions. People usually experience stochastic events in ways that interfere with learning normative statistics, whereas a videogame could provide a helpful set of experiences. This claim was strongly supported. Both the game and the passage had significant and independent positive effects on learning. The strongest demonstration was that compared to the baseline of no game and no passage, students who completed the game in distribution mode and read the passage exhibited a two-standard-deviation learning advantage.

The more theoretically interesting question is whether playing the game would improve subsequent learning from the passage compared to just reading the passage. This received reasonable but not definitive support. The students who played the distribution-mode game and read the passage did better than students who only read the passage. The significance of this difference depended on a one-tailed test, so the stability of the finding is still somewhat tentative. On the other hand, in terms of effect size, the distribution-mode game plus passage yielded a large learning gain of .84 standard deviations more than the passage-only condition.

A second major hypothesis was that videogames may not show learning effects on standard SPS measures and that therefore PFL measures are more appropriate. The results supported this claim and showed that the value of the passage was amplified by playing the game beforehand. The advantage of PFL measures over SPS measures was best demonstrated by the results of the proportion-mode

version of the game. Without the passage, the proportion-mode version of the game appeared no better than the no-game/no-passage control condition on the posttest; but when proportion-mode students received the passage, their learning gains were quite large.

Surprisingly, the distribution-mode game did not show the same pattern. Participants who only played the distribution-mode game did better than the control students, even without the passage. One possible explanation is that this version of the game used distributions as part of the game display, so participants were better able to answer the test questions that used images of distributions, even without subsequent formal instruction. In other words, as one might expect, it is still possible for a game to include elements that can be detected by sequestered-problem-solving measures. We would not want to claim otherwise. But the distribution-mode finding does not undermine the key point demonstrated by the proportion-mode game condition, which is that the value of experiential learning is better captured by PFL measures than by SPS measures.

Our final hypothesis was that given the passage, the distribution version of the game would yield better overall learning than the proportion version of the game. This was not the case. The distribution-mode game did fare better, but the difference was far from statistical significance. This was a surprise, because we had theorized that including a representational form that helps organize instances according to their probabilities would provide a way for students to encode their experiences in a more normative manner. An *ad hoc* account for the lack of effect is that the bottom of the screen (Fig. 1d) provided a histogram of the individual alien invaders that the player destroyed, effectively creating a useful representation for the proportion-mode students. Regardless, the current research did not provide support for the importance of including well-organized representations as part of one’s experience in preparation for learning from a subsequent explanation.

## Conclusions

These study results are consistent with the hypothesis that playing *Stats Invaders!* gave participants intuitions about the behavior of probability distributions that, in turn, prepared them to learn from subsequent formal instruction on the topic. This is important because probability is a notoriously difficult topic about which to form normative intuitions (Tversky and Kahneman 1974), and videogames provide a new approach. In our best model of instruction (distribution mode plus a passage), students correctly answered about 64 % of the questions at posttest, up from 34 % at pretest. This is not perfect accuracy, so there is still more to be done to improve learning. At the same time, the

percentages indicate that our assessment and game addressed a difficult domain of learning with some success.

Of more central importance, the research demonstrates that even without having instructional content that maps directly onto curricular standards, games can prepare students to learn in more formal environments, such as school. Game environments can provide experience, and formal environments can provide explanations. This clarification should be useful for those creating learning games, because it can help them focus on what games do well, rather than trying to make games into a stand alone solution for learning. This latter alternative can lead to sugar-coated drill and practice: game designs that rely on the motivational powers of points, levels, graphics, and narrative, but give up the experiential potential because designers feel pressure to deliver declarative and procedural content.

A common observation about experiential training, ranging from digital simulations to interpersonal role-play, is that the learning benefit “really” shows up during debriefing afterward. The implicit insight of this observation is that directly measuring the learning value of a game may be misleading when the goal is to deliver compelling experiences. In this scenario, a more appropriate measure is to determine how well the game prepares students for future explanations, whether delivered through discussion, lectures, or text. This does not mean that designers of learning games are liberated from the shared responsibility of ensuring that students learn. Rather, it liberates game designers from the belief that their game is successful only if it works by current sequestered achievement measures. It may be more appropriate to measure whether the game prepares students for future learning. We suspect that many compelling games have been “shelved” for educational purposes because their effects were mismeasured.

Of course, not every videogame is going to prepare students for subsequent explanations. In our case, we built *Stats Invaders!* using theories about learning in general and statistics specifically, and we borrowed a successful play pattern. We did this to test a hypothesis rather than to create a template for game design. However, we used a generalizable process that seems worthwhile for design teams. If the ultimate goal is to help students learn explanations, as opposed for example to building automaticity, then a key move in the design process is to determine what experiences people need to make that explanation meaningful. A second key move relevant to game design is to decide how to help people lift out the important elements of those experiences for future learning: for example, by creating a core mechanic that emphasizes the way experts organize their experiences.

**Acknowledgments** This material is based upon work supported by the National Science Foundation (Grant No. SBE-0354453) and the

MacArthur Foundation’s Digital Media and Learning Initiative. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the foundations. The authors thank Don O’Brien for helping with the programming of *Stats Invaders!* and the best interface touches.

## References

- Anderson D, Lucas KB, Ginns IS, Dierking LD (2000) Development of knowledge about electricity and magnetism during a visit to a science museum and related post-visit activities. *Sci Educ* 84(5):658–679
- Arena DA (2012) Commercial video games as preparation for future learning. Unpublished doctoral dissertation. Stanford University, Palo Alto, CA
- Arena D, Schwartz DL (2010) Stats Invaders! Learning about statistics by playing a classic video game. In: Proceedings of the fifth international conference on foundations of digital games, ACM, New York, pp 248–249
- Bandura A, Ross D, Ross SA (1961) Transmission of aggression through imitation of aggressive models. *J Abnorm Soc Psychol* 63(3):575–582
- Barron B, Wise S, Martin CK (2012) Creating within and across life spaces: the role of a computer clubhouse in a child’s learning ecology. In: Bevan B, Bell P, Stevens R, Razfar A (eds) *LOST opportunities: learning in out-of-school-time*. Kluwer, Netherlands, pp 99–118
- Bransford JD, Schwartz DL (1999) Rethinking transfer: a simple proposal with multiple implications. *Rev Res Educ* 24:61–101
- Edwards W (1968) Conservatism in human information processing. In: Kleinmuntz B (ed) *Formal representation of human judgment*. Wiley, New York, pp 17–52
- Faul F, Erdfelder E, Buchner A, Lang A-G (2009) Statistical power analyses using G\*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods* 41:1149–1160
- Feuerstein R (1979) *The dynamic assessment of retarded performers: the learning potential assessment device, theory, instruments, and techniques*. University Park Press, Baltimore, MD
- Fox J, Weisberg S (2011) *An R companion to applied regression*, 2nd edn. Sage, Thousand Oaks, CA
- Himmelberger K, Schwartz DL (2007) It’s a homerun! Using mathematical discourse to support the learning of statistics. *Math Teach* 101(4):250–256
- Ito M, Gutiérrez K, Livingstone S, Penuel W, Rhodes J, Salen K et al (2012) *Connected learning: an agenda for research and design*. The MacArthur Foundation, Chicago
- Konold C (1989) Informal conceptions of probability. *Cogn Instr* 6:59–98
- Konold C (2007) Designing a data tool for learners. In: Lovett M, Shah P (eds) *Thinking with data*. Taylor & Francis, New York, pp 267–291
- L’Ecuyer P, Meliani L, Vaucher J (2002) SSJ: A framework for stochastic simulation in Java. In: Proceedings of the 34th conference on winter simulation: exploring new frontiers, IEEE Press, pp 234–242
- Lehrer R, Kim M-J, Ayers E, Wilson M (2013) Toward establishing a learning progression to support the development of statistical reasoning. In: Confrey J, Maloney A (eds) *Learning over time: learning trajectories in mathematics education*. Information Age, Charlotte, NC
- National Research Council (2011) *Learning science through computer games and simulations*. In: Committee on Science Learning: Computer Games, Simulations and Education, Honey MA,

- Hilton ML (eds), Board on Science Education, Division of Behavioral and Social Sciences and Education. The National Academies Press, Washington, DC
- Nisbett RE, Krantz D, Jepson C, Kunda Z (1983) The use of statistical heuristics in everyday inductive reasoning. *Psychol Rev* 90:339–363
- Petrich M, Wilkinson K, Bevan B (2013) It looks like fun, but are they learning? In: Honey M, Kanter D (eds) *Design, make, play: growing the next generation of STEM innovators*. Routledge, New York
- Schwartz DL, Bransford JD (1998) A time for telling. *Cogn Instr* 16(4):475–5223. doi:[10.1207/s1532690xci1604\\_4](https://doi.org/10.1207/s1532690xci1604_4)
- Schwartz DL, Martin T (2004) Inventing to prepare for learning: the hidden efficiency of original student production in statistics instruction. *Cogn Instr* 22:129–184
- Skinner BF (1986) Programmed instruction revisited. *Phi Delta Kappan* 68(2):103–110
- Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185:1124–1131