Title
Designing Games for Understanding in a Data Analysis Environment

Journal
Technology Innovations in Statistics Education, 7(2)

Author
Erickson, Tim

Publication Date
2013

DOI
10.5070/T572013897

Designing Games for Understanding in a Data Analysis Environment

1. INTRODUCTION

Data Games is a project funded by the US National Science Foundation that has developed web-based games embedded in a data-analysis environment. As students play, the system collects the data, and students can analyze it. The games are designed so that data analysis is rewarding, that is, it's the best way to improve your performance. Put another way, these games can also be a constructivist platform (Jong, Shang, & Lee 2010) in that students bring mathematics to bear to accomplish their own goals, namely, winning.

Our cadre of 9 field-test teachers were able to use these simple Data Games in regular classes; and when they did, almost all students were engaged, and appeared to learn and succeed. Data Games appear to give students and teachers an engaging, accessible, and effective introduction to data science and modeling.

Data Games do not need to be about statistics. Rather, they are an engaging environment in which "regular" math is useful. The skills and habits of mind of data science offer students a rich, active approach to mathematical understanding. A data-oriented approach like this can reinforce concepts and skills; help students see how principles apply in many contexts; and for some students, unlock math topics that have eluded them in the past.

To make this work, however, one needs suitable games.

This paper begins with a discussion about designing these games for learning, gives examples of Data Games, and reflects on field-testing in mathematics classes in San Francisco, California high schools (ages 16–18). We postpone discussion of theory and research until after these examples; then we will see where these games fit in the extensive work being done in modeling in mathematics education. The final section of this paper will draw conclusions and make suggestions for further development and investigation.

1.1 The Big Issues of Data Game Design

Students want to do well in games. If data analysis helps them win, they will want to learn how to analyze the game data. Ideally, this experience in data analysis transfers to data analysis outside the game context. This leads to some important requirements:

- Data analysis has to be *useful*. Analysis must genuinely help the player succeed in the game.
- Data analysis has to be *easy*. If it's hard, it won't be worth the effort.

Making data analysis useful is where the "art" lies in the game design. But the data analysis takes place outside the game itself. The games are embedded in a web-based data-analysis environment (currently called "DG") so that the data collection is automatic and so that it is easy to make graphs and do calculations. Also, the activities themselves—the student instructions, for example—are designed and presented separate from the games. That is, the games do not teach data analysis: they are an environment in which data analysis is useful.

Beyond Data Games, there are many rich settings where students can be data scientists and where data analysis plays a central role. These settings range from real-life projects to analyses of complex single-player games to multiplayer online games. We could even structure them as "Model-Eliciting Activities" (Lesh et al., 2005). Yet few secondary-school teachers have the time or expertise to facilitate such open-ended student experiences. This leads to a design decision:

- These games have to be *short*. The play and the data analysis have to happen in minutes, not hours of play.

These three design principles—that the games and the analysis must be useful, easy, and short—imply that these games should be small-scale and simple. It was easy to imagine interesting complications, extensions, and rich contexts. But in classroom tests, it always turned out that the simple version was better.

These three principles will become clearer with examples. But first, a few words about the students for whom these games are intended.

1.2 The Target Audience and Appropriate Topics

Data Games are intended for students aged approximately 12–18. The Data Games that appear in this paper integrate data science into the mathematics curriculum. Although they are suitable for a statistics classroom because they involve data, they focus on content that many statistics curricula assume students have already mastered.

Many of these students in our field-test classes had not been successful in secondary mathematics. Even though many had several years of secondary math instruction, they had trouble finding meaning in graphs and understanding how multiple representations of data correspond to one another.

As a consequence, the teachers and students spent substantial time reinforcing important mathematical concepts and skills, for example, fractions, formulas, direct proportion, the meaning of slope, and reading graphs. Although these are not traditional topics in a statistics course—except perhaps in the first week—they are the foundation for many basic statistical ideas as well as essential understandings for secondary mathematics (Ben-Zvi and Garfield 2004).

So this paper begins by focusing on topics like these, partly because they were so useful for these students, but also because they show how we can teach these topics in a data-rich manner.

These games are useful for teaching statistical topics as well, for example, coping with variability. We will discuss variability and other statistical issues later in the paper.

2. "PROXIMITY" DESIGN DISCUSSION

This section explores how these game-design principles and decisions relate to a game called *Proximity*, and how field-testing positively impacted the specifics of *Proximity's* design.

2.1 Example Game: *Proximity*

In *Proximity*, students try to propel a white ball to the middle of a target. The closer the ball is to the center of the target, the higher the score. A "bullseye" is worth 100 points. There are 6 balls per game, and the target moves after each shot. Players can use an on-screen ruler to measure distances.

Figure 1. The game *Proximity*. The user tries to get the white ball to the center of the yellow target. In the illustration, the "push" is 29.3.

Students learn to "drive" the interface quickly, and seem to intuit the object of the game. If we ask a student, "how did you know how hard to push the ball?" they know that to go farther they have to push harder. But their understanding is not quantitative; we encourage them to make a graph. With the click of a button, they can make a scatter plot like the one in Figure 2.

Students can easily put a movable line on the graph and find its slope. (They can even lock its intercept to go through the origin if they have that insight.) If they measure the distance before they shoot, they can divide by the slope to figure out how hard to push the ball.

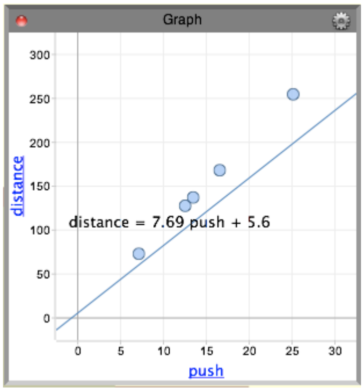

Figure 2. A graph from *Proximity* showing the distance the ball goes as a function of push. A "movable line" is on the graph; its equation changes as the user moves the line.

Just playing "by feel," students can get scores over 250 without much trouble. Using the graph and the formula for the line, they can beat 450 reliably.

Thus data analysis improves performance. It's easy because the data are automatically collected in the DG environment. It's a quick game with a short learning curve: six balls and you're done. Students can get to a computer lab, play many games, analyze, debrief, and be back in an hour.

2.2. Student Behavior with Early Versions of *Proximity*

But the fact that students *can* improve their scores through data analysis does not mean that they *will*. In general, despite evident interest in the games and enthusiasm for being at the computer, field-test students playing prototype Data Games often failed to use the tools they had available. For example, we often saw students write down—on paper—data that were displayed in a data table on the screen. More significantly, beginning students seldom made graphs to show patterns in their data; and when they made a graph, and used a movable line to model their data, they seldom used the equation that appeared on the screen in their calculations.

These students were not being contrary or lazy. Here are some conjectures for why they behaved as they did: Without technology, you have to write everything down yourself, and graphing is hard. These students were not used to using technological tools. Therefore, they fell back on the familiar: they recorded numbers and looked for patterns primarily in tables rather than in graphs.

Put another way, there is an "activation energy" to using a tool like a graph; if students have a familiar technique (looking for patterns in a table of numbers) they'll prefer it even if the new one (a graph and its formula) would save them time, confusion, and agony. This "tool aversion" may be particularly acute when formulas are involved. Even though students do not need to come up with the formula themselves—in this case, the movable line gives it to them—actually using that formula in order to accomplish something is unfamiliar and may be uncomfortable.

This problem is made more difficult by habits students seem to have developed playing recreational games. Students easily fall into a pattern of pressing keys—filling in numbers and pressing return—as rapidly as possible. The (un)reasoning seems to be: if you can get a high score by playing five games mindlessly in the same time it would take you to play one game thoughtfully, you should go for mindlessness. It is therefore a good idea to slow the games down, or otherwise make rapid play unrewarding, so there is a greater benefit to being thoughtful.

But fundamentally, we think students are not used to using data and models anywhere in their problem-solving process. We want to help them develop habits of mind that make this easier. The next section describes part of the trajectory of *Proximity*'s design, and some of the ways we altered the game to make more rewarding for students to use the tools and develop data-aware habits of mind. In a later section on assessment, we'll see how we can tell whether students actually develop these habits.

2.3. Making the Data Essential

It is hard to design a game where data analysis is *useful*. If students can succeed in the game "by feel"—without using the data, and without much mathematical understanding—it has failed as a data game.

In the first level of *Proximity*, for example, students quickly discover that the distance the ball travels is 10 times the "push" they give the ball. So if students succeed, we know only that they can measure distance and divide by 10. Given a little experience—and class chatter—everyone can do well without really using data to determine a strategy. Thus the first level is a good introduction to the game, but success at that level is not evidence that students can come up with a winning procedure from scratch.

That requires a second level, where the slope—the coefficient in the distance-push relationship—is different every game, that is, every set of six balls. Students can't simply divide by 10; they have to look at each game anew and deal with its data.

Making this work is not trivial, however. It took repeated classroom field-test sessions to see what students actually do. Simply making the slope vary was not enough; at first, students reverted to playing by feel. Our response was to create a system of requirements, so students had to achieve a particular score on a level in order to move up to the next. That way, a lesson could (for example) challenge students to pass through the second level and get to the third.

Back in the classroom, this proved to be too onerous; too often, the ball would get stuck near an edge, and it was too hard to hit the target. Students tried to use the data but became discouraged. Lowering the required score, however, invited playing by feel again. Therefore, in the next version, the interesting and challenging edge shots (alas) no longer appeared—and the score requirement increased.

Still, some students would get lucky, and by chance, the slope they had to figure out was near what it had been before; or they just guessed well. So getting a high score was mechanical: measure the distance, divide by some number. They didn't need to look critically at their data. So in the next version, you had to beat a score (425) twice in a row to advance. The software guaranteed that the slopes would be substantially different from 10 and from one another.

(Why 425? Zero push gives zero distance, so line must pass through the origin. Since this level is completely deterministic—no variability—the first ball determines the slope; if it misses completely and gives no points, an average of 85 on each remaining ball achieves the goal of 425 points. That seems the right level of difficulty.)

This whole dance may seem obvious in retrospect, but it points out the vital importance of field test.

2.4. Skill and Luck

Now consider the balance between skill and luck. Mathematical skill is of prime importance, but luck plays two important roles. The first is when the mathematical lesson involves probability. One of the games, Wheel, is a roulette-like betting game designed to teach students about the law of large numbers and expected value. So some students will be lucky and win in the short term, even without any data analysis; but they will lose in the long run. (Patient students who analyze the data well can uncover a winning strategy, however.)

Luck's second role is to create interest; if you don't know exactly what's going to happen, you're more likely to pay attention. (This can frustrate more advanced students: despite understanding everything perfectly, they cannot get a perfect score. But these students are the exception—and not the population of our greatest concern.)

Luck is linked to variability. If everything is completely determined, it's uninteresting. Furthermore, we want students to learn to cope with variability. Variability comes in many forms in a data game. In *Proximity*, for example, there are several sources:

- At higher levels, the relationship between push and distance has a little randomness in it.
- Not all "push" values are possible. As you prepare a shot, the mouse pointer's position is quantized to the nearest pixel; the internal relationship between those pixels and the push means that the value jumps in discrete steps. You can always get close, but you seldom get the precise value you want.
- There is variability in how a student fits a model to the data—not everyone chooses the same movable line—so calculations based on models will give varying results.
- There is variability in how well players aim at the target; if the aim is off by a few degrees, the ball will miss the center of the target and will not yield the highest possible score.

This last source is particularly interesting. It has at least two drawbacks, though: first, it rewards game-playing experience, which we want to avoid (likewise, none of our games have a "twitch factor" that rewards fast reactions). Second, and more subtly, if students got a lower score simply because of poor aim, we would have to lower the threshold for success—and it would be easier for someone who does not really understand the mathematical model to succeed.

In the current design, based on field test experience, if a student points close to the direction of the target, he or she will automatically point in the right direction. A student who makes a good model and understands it can get a vey high score—subject to the pixel-quantization problem mentioned earlier.

This is an example of a long-standing tension in designing technology for learning: when to make the student do things themselves, and when to step in and do it for them? In this case, aiming perfectly is not an important learning outcome, so the game takes care of it—in a way that many users will not even notice.

The next sections explore design issues that apply to other games as well as to *Proximity*.

3. CURRICULUM SURROUNDING THE GAMES

A lesson based on a game is more than simply playing the game. Lessons that seem to work best have three phases:

- an introduction;
- a time for play and basic data analysis;
- some sort of consolidation phase.

For the introduction, it is usually sufficient to tell the students to play. If the game is well-designed, they will learn the game mechanics automatically, uncovering the few subtleties with a little practice and exploration. In addition, animations and short instructional videos introduce each game. This introductory phase also gives students a chance to get some of the initial distraction of play out of their systems.

This phase can take place in the classroom, or can be assigned as homework. For a lesson using *Proximity*, homework might involve requiring students to score 300 or better, or to unlock the second level; in order to do that, they have to have learned how to shoot the ball at the target and possibly to measure distances.

The second phase challenges students to use data analysis to improve their performance. Put another way, students transition from simply playing the game to using math as they play. Students do not generally do this on their own; this requires some direction from the teacher or the video, and some help using graphs and other analysis tools. In *Proximity*, for example, some students need to be prodded to look at the graph and figure out an appropriate push based on the equation of the line. Often a question like, "how did you decide to use that amount of push?" will help students move in the right direction.

In the classroom, this generally takes the form of students working in pairs to accomplish some goal such as unlocking a higher level. Hard-copy or online worksheets can accompany the game play, asking questions that probe both the basics of the game mechanics ("What's the maximum score you can get in this game?") and more sophisticated issues ("Describe how you use the graph to decide how hard to push the ball.").

After students have done this, they need to consolidate and solidify their learning. This is the third phase. Although we can ask students to reflect, or write about what they learned, we like a more performance-based approach. What could students do to make their learning more explicit and apparent?

One intriguing answer made possible by technology is to have students *automate* the process of playing the game. That is, create instructions so that a "robot" can play and win. This has two advantages:

- It takes the students out of simply playing and forces them to be explicit about their strategies. Furthermore, this often involves encoding their ideas in symbolic mathematics.
- It relieves the tedium of an artificially-slow game. Automated play is *faster*, which seems to be a reward to students.

There are design challenges here: how do you get students to teach a robot to play without programming? We're working on that; one answer is to restrict the types of strategies students can employ. In low levels of *Proximity*, for example, where the results are deterministic, the first push could be a small fixed value, which can be used to completely determine the slope. Then on balls 2 through 6, the robot could use the that slope, and the distance to the current goal, to determine the push value for each ball. This requires students to enter a single formula rather than writing a more general program. (In one field test, a pair of ingenious year-7 students decided to start every game with a "push" of 1.)

Player strategies have been most successfully implemented in the game *Markov*—our version of rock-paper-scissors—where there are only three choices available for every move, so students simply push buttons to specify their choices for each of the nine possible game situations.

4. ASSESSMENT AND DATA GAMES: THE LOGS

If students play games, how do we know they learn anything? Ordinary assessment practices work fine; students can perform tasks to show what they understand about the math topics in the game. These tasks can be proximal, i.e., framed in the context of the game, or distal: more concerned with the general topic (Ruiz-Primo et al., 2002).

But because the games are played on the Web, the software can record everything the students do. The system records every move, every game and score in a game log. These logs tell us when students unlock new levels, when they make graphs, the equations of the lines students put on graphs, and the calculations they make with the in-game calculator. This means we can study game play in detail. It's a new window into student thinking.

We'll see how this works by discussing a different game, *Cart Weight*.

In *Cart Weight*, the player guesses the weights of five carts that have different numbers of bricks on them. In the first level, "Dubuque," the cart is weightless and each brick weighs 3 units. Most students see the pattern immediately. In the second level, "Ames" (shown in Figure 3 below), the cart weighs 8 and each brick weighs 4, so students have to deal with the intercept.

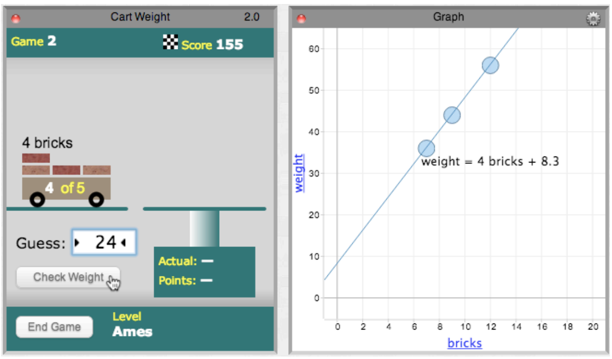

Figure 3. The game *Cart Weight* and a graph showing the weight of three previous carts as a function of the number of bricks. The user guesses that the current cart, with four bricks, weighs 24 units.

In the "Davenport" level, however, the slope and intercept change every game. So students have to create their models thoughtfully and efficiently; they can't easily succeed without a good graph, and a line to model the weight data.

Figure 4 displays log data. It shows a session from a single pair of students working for 20 minutes. Each point is an event; events are described in the legend. For example, the tall green rectangles appear whenever a student adjusts a movable line on the graph; so we can tell that the students didn't really use the graph until the tenth minute, but then used it extensively:

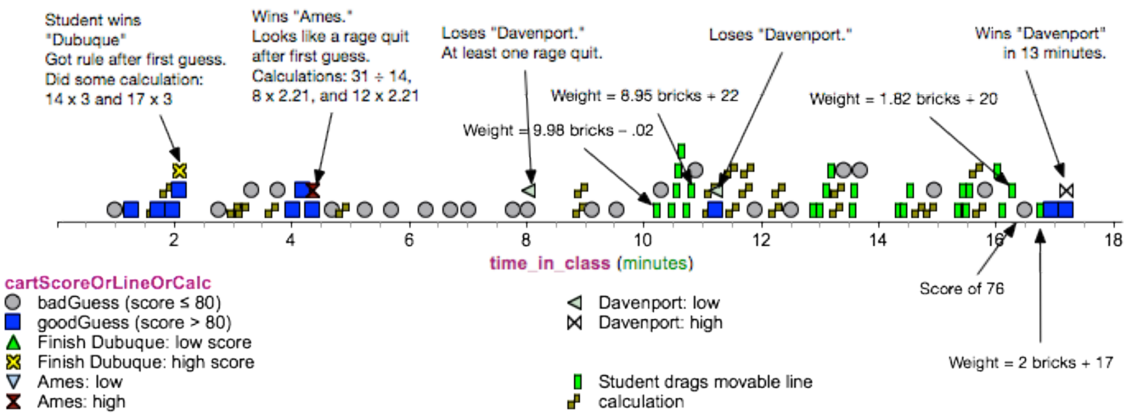

Figure 4. A visualization of a log showing student work on *Cart Weight*. The horizontal axis is in minutes.

This log data lets us reconstruct the session: The students mastered "Dubuque"—the first level—in two minutes. (Note the legend: the hollow "X" indicates a high score on Dubuque.) It took two minutes more to master "Ames." But then there were many poor guesses (gray circles, indicating a score of 80 or below on the guess) as the students puzzled out the third level, "Davenport" (minutes 5–10, making a few calculations). Finally, in minute 10, they made a graph, and the log shows that they put a line on the graph and adjusted it repeatedly. Just after that, they made their first good guess in Davenport (square at minute 11)—but not good enough to get a high score for the whole game.

Between 12 and 14 minutes the students made many calculations and adjustments, but still made four bad guesses (gray circles). Realizing they had a low score, they did a "rage quit" and started a new game before that game ended. At minutes 15 and 16, they made the first two guesses for the next game—poor guesses, but that's OK: they needed two points to determine the line. At 16:30, they made a line with the equation **Weight = 1.82 * bricks + 20**, and got a pretty good score of 76 on the next guess. They then adjusted the line to **Weight = 2 * bricks + 17**, and aced the last two carts, giving them a high score for the level.

This period of time—a scant 20 minutes—constitutes the first phase and most of the second of an activity about linear relationships. *Cart Weight* has no "automation" mode yet, so the third, consolidation phase was a paper handout where, among other things, students wrote about the meaning of the numbers in the equation for the line (the slope is the weight of a single brick; the intercept is the weight of the cart) and wrote instructions for how to beat the Davenport level.

These logs give us an intriguing window into a student's thinking. The logs show early success, then struggle, then bringing analytic tools to bear, and finally success on a more challenging task. We hope to connect logs like this one with other artifacts and observations, and with logs from later sessions. This can help assess the elusive "habits of mind" we would like nascent data scientists to develop. For example, do students turn to graphs more quickly in subsequent sessions? The logs can answer that question.

5. DATA GAMES AS MODEL-ELICITING ACTIVITIES

The Data Games project facilitates teaching modeling, a concept and skill of great importance to science, mathematics, and statistics education. There are many definitions for modeling; here's one:

> A model is an abstract, simplified, and idealized representation of a real object, a system of relations, or an evolutionary process, within a description of reality. (Henry, 2001, p. 151; quoted in Chaput et al., 2008)

Thus a model can be a probability model for some stochastic process; a function describing quantitative data from a phenomenon; a set of linked differential equations describing a dynamical system; a Markov process; a rating system for making a decision; or even a simple diagram. The key element is abstraction: models simplify reality and make it tractable; the test of a model is how well it matches the aspects of reality we're concerned with. Whence George Box's (1987) famous quote: "All models are wrong, but some are useful."

Garfield et al. (2012) make a good case for better student learning through modeling activities; but modeling extends well beyond statistics education. Modeling materials have appeared under the aegis of applied mathematics (e.g., Erickson 2005; Engel 2010). And models in introductory physics closely resemble the ones in this paper: they're functions on the Cartesian plane, fit to data points. Research on modeling in high-school physics is extensive and positive (e.g., Jackson et al, 2008), and also focuses on functions as models for phenomena.

Our design of Data Games—the games themselves and the lessons that surround them—follows the overarching principles laid down by Garfield et al. (2012) in their design of the CATALST materials. Naturally, our lessons for secondary school are at a very different level and scope than a statistics unit for university students, but the direction is the same: use technology to help students do meaningful modeling (e.g., Doerr and English 2003), and through that, help them build deeper understanding.

5.1. Model-Eliciting Activities

What makes a good modeling curriculum? Garfield specifically mentions model-eliciting activities (MEAs) as described by Dick Lesh and his colleagues (e.g., Lesh, et al. 2000). These MEAs are rich, large-scale, open-ended activities, usually accomplished by groups, that pose real-world problems in which the students construct and test models as part of their problem-solving process. In addition, the work is more than simply an answer: it's elaborate enough to reveal student thinking. One can think of Data Games as "entry-level" MEAs, suitable for students or teachers who do not have the background, skills, or time for the full-blown version. The following list examines how Data Games map onto six principles Lesh lists as essential for model-eliciting activities:

- Model-construction principle. (Problems must allow for the creation of a model.) Students win at Data Games by constructing good models. Students quantify and predict to make their moves, and the game itself gives them feedback. But the games don't explicitly force students to reflect and explain; so although this is a start, it is not as sophisticated as in tertiary-level MEAs. Still, we can ask students to reflect, and we can use the game logs to see out what models students constructed, and how students used them.
- Reality principle. (Problems must be meaningful and relevant to the students, ideally to their everyday lives.) The data are real and meaningful, but only in the context of the game, not in students' everyday lives.
- Self-assessment principle. (Students must be able to self-assess, or measure the usefulness of their solutions.) Students generally do not assess themselves (though this does happen in the *Lunar Lander* game, not discussed here). Instead, students' experience in the game lets them know how well their model worked—but not, for example, whether their procedure was optimal.
- Model documentation principle. (Students must be able to reveal and document their thinking processes within their solution.) This is present in the debriefing, "consolidation" phase of the lesson. This is much less extensive than in a "real" MEA, but Data Games lessons do ask students to explain their models and their consequences (e.g., limiting cases), and the meanings of any parameters, if possible.
- Model re-usability principle. (Solutions created by students are generalizable or easily adapted to other similar situations, rather than limited to a specific problem.) Data Games

models are adaptable to other games and game-problems, particularly as one progresses to higher and higher levels. See the next section for more on this point.

- Effective Prototype Principle. (The model should be as simple as possible, but still mathematically significant. It should provide a useful "prototype" for interpreting problems with the same underlying structure.) Many games are designed precisely to be prototypes for a class of situation. *Proximity*, for example, is a situation involving direct proportion, $y = kx$, in which you know y and k, and need to find x. This is a big idea in understanding quantitative relationships, so students and teachers can refer back to the game as parallel situations arise.

5.2. How Thinking About MEAs Helps: Models, Levels, Graph Resistance, and Re-Usability

The re-usability principle in MEAs asks that whatever model you make be applicable to other situations. To illustrate re-usability in Data Games, and to look more deeply at the modeling students do, we look again at *Cart Weight*, the game in the "Logs" section above.

In the first level, the weight of the cart is three times the number of bricks. Students notice this relationship quickly, and easily get enough points to advance. They also tend to use only calculation, but do not refer to the graph; they do not use movable lines or their equations as tools. And there is no need to: students have developed a model, but not one sophisticated enough for MEAs. It only applies to this one situation. It does not generalize.

In the next level, the cart weighs 8 and each brick weighs 4. The numbers are easy to deal with. Some students figure it out, and the answer flits around the room. Again, students do not need the line or the equation. And again, they have a model—but its use does not extend beyond this level of the game.

In the third level, the cart's and bricks' weights change from game to game. And the scoring is arranged so that, in order to master the level, students must predict the weight very precisely for at least three of the five carts. They must either: put a line on the graph and use the equation; solve a system of two equations and two unknowns; or use a combination of intuition, guesswork, and brute-force reasoning to solve the problem. This is the level at which the relevant modeling tool—the line on the scatter plot with its accompanying equation—finally becomes the path of least resistance. Not coincidentally, this is also the point at which the procedure—use two points to determine the line, then use the equation for the line to predict the weight—becomes reusable.

So according to Lesh, Cart Weight doesn't demand serious modeling until the third level. The first two levels only familiarize students with the setting; and the log data above shows that in that group's case the initial levels were effective in that role. Those students created a model and used it to succeed. But students who never get to the third level haven't really grappled with modeling.

This careful design and balancing that encourages students to use that equation reveals an emergent design principle for Data Games lessons. In order to induce students to use modeling principles and tools, the games must:

- Erect barriers to success when you don't use the tools, e.g., by requiring very high scores to advance; and
- Lower barriers to using the tools themselves, e.g., by making them easy to use and giving suggestions to make a line and use the equation.

Doerr and English (2003) point out how important it is for students to have multiple exposures to modeling experiences. Another aspect of re-usability and the prototype principle in Data Games is the way the games reinforce one another: *Proximity* and *Cart Weight* (and a third game, *Shuffleboard*) all have different "takes" on linear functions, and all use the movable line as the technological tool to get the model for the data.

A side note: On the fourth level of *Cart Weight*, the weights are no longer integers. In field tests, the moment the first decimal weight appeared on the screen, students audibly gasped. “Oh no, decimals,” was the near-universal reaction—despite the fact that the procedure for finding a model would be identical to that in the third level. Here is a conjecture: with integers, students still had the impression that they could solve the problem in their heads, “without math.” Integers are comfortable and manageable. They could solve it by trial and error. But decimals demand lines, formulas, and calculation, and the students were uncomfortable.

6. DATA GAMES FOR STATISTICS

This paper has focused on using Data Games for what is essentially *mathematics* instruction as opposed to *statistics* instruction. But teachers can use Data Games for statistics as well. After all, the games are just sources of data; we can create game situations in which understanding statistical ideas improves performance in the game.

We alluded above to Wheel, a wheel-of-fortune sort of game, where observing for long enough will reveal the biases in the wheel, and an understanding of expected value will show you the way to a moneymaking strategy.

Even more of these have been created by the Data Games group in Amherst, MA, USA, especially for middle-grades students, and address issues such as understanding variability (e.g., Shaughnessy et al, 2004). These include *Ship Odyssey*, a treasure-hunt game where players get imperfect information about the location of a treasure. Another game, *Rock Roll*, explores issues behind experimental design. One can imagine endless variations, for example, a game where stratified sampling gives better results than a SRS. The art will come in deciding exactly what the point of such a game is, and tuning the game so that an understanding of sampling really does make a difference.

Variability can make a game more interesting. What role does variability play if we consider these games in the context of statistics education?

Wheel is straightforward: it’s a gambling game, and there’s randomness. The variability is central and authentic.

Higher levels of *Proximity* include a little noise to the distance a ball travels. Students have to cope with the notion that the line may not go through all the points as it did in lower levels (and as it generally does in math textbooks). It also means that the strategy of using the first point to compute the slope is insufficient; a data-aware strategy would be to recompute the slope each time to be the average distance divided by the average push.

In *Ship Odyssey*, there’s an underlying story: players send specially-trained rats down to find the treasure. After they find the treasure, they swim back to the surface. Alas, the turbulent waters force them off course in a random fashion, creating a distribution of rat-surfacing locations; players use those to decide where to drop their grappling hook. That is, the “rat” feature in the story introduces noise. The advantage is that it makes the story is fun and engaging; the disadvantage is that it goes a bit against our drive to simplicity: the rat narrative may get in the way of understanding the data—at the very least, it takes time to explain.

Another game with statistics content is *Floyd’s of Fargo*. This is an insurance game; the player is insuring cars against flat tires (as in the rat narrative, the context consumes class time). The player must set the premium price. A new tire costs $100. The lower the premium, the more people will buy your insurance—but the more you’ll have to pay out in claims. The object of the game is to make as much money as possible.

There is an optimum price, and students can find it theoretically. The number of customers depends on the price, and the number of flat tires that must be repaired depends on the number of customers. It’s a matter of a little algebra (and understanding of expected value) to find the equation for the quadratic that represents the profit as a function of premium price.

But empirically—that is, when you are actually playing the game—things are not that simple. The number of flats experienced by the customers will seldom be the expected number: it's a binomial random process. So as an insurance company, a player's profit in any given turn varies considerably around the expected value.

Thus students must cope with determining a model in the face of randomness—in this case, randomness that arises naturally from the context.

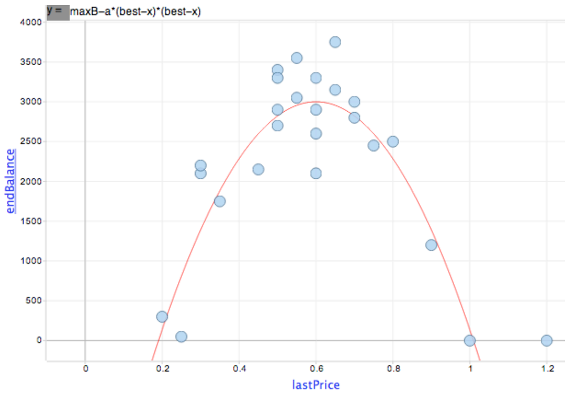

Figure 5. Ending balance in *Floyd's of Fargo* as a function of premium price. The graph shows a model quadratic, in the process of being fit. Its formula is in vertex form (above the graph). The parameters *cust*, *a*, and *best* are controlled by sliders.

Students can therefore play this game on many levels. It can be a modeling game, where they find the appropriate graph, fit a parabola, and realize which parameter gives them the largest profit. Or it can be a rich optimization problem. They can also study the nature of the underlying probability model and see how the number of flats varies given the same number of customers.

Thus statistical issues appear in these games in different ways, through the deliberate introduction of noise or through the choice of phenomena where sampling and combined events create variability naturally. All of these present students with data that vary, very much like data they will encounter in real life.

One could develop Data Games purposefully to address specific topics in statistics education, for example, in the seven areas of variability compiled and elucidated by Garfield and Ben-Zvi (2005). We could use the learning analytics tools described above—instead of or in addition to test items—to assess student understanding of these in a "performance" context.

7. CONCLUSION AND DIRECTION

Playing a Data Game (like most instruction) does not create understanding. But Data Games and the associated activities are a good opportunity to develop and solidify understanding of mathematical ideas. Likewise, success at a Data Game is not ironclad evidence of understanding or of a well-engrained habit of mind, but it can be a useful tool in a teacher's assessment arsenal.

The logs are a particularly exciting windfall from this project. We have never before been able to see so many students' progress in such detail. Would access to them in real time help us make valuable instructional decisions on the fly? That's one direction for future work.

Then there's the question of the breadth of topics these small games might encompass. We encourage readers' ideas and participation as we come to understand what's possible and move forward to design new games and activities.

Avid readers interested in the project should visit http://play.ccssgames.com.

REFERENCES

Ben-Zvi, D. & Garfield, J. eds. (2004). *The challenge of developing statistical literacy, reasoning, and thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Box, G. E. P. and Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces*, New York: Wiley.

Chaput, B., Girard, J., & Henry, M. (2008). Modeling and simulation in statistics education. *Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*.

Doerr, H. & English, L. (2003). A modeling perspective on students' mathematical reasoning about data. *Journal for Research in Mathematics Education*, 34(2), 110–136.

Engel. J. (2010). *Anwendungsorientierte Mathematik: Von Daten zur Funktion*. Heidelberg: Springer.

Erickson, T. (2008). A Pretty Good Fit. *The Mathematics Teacher*, 102(4), 256–262.

Erickson, T. (2005). Stealing from Physics: Modeling with Mathematical Functions in Data-Rich Contexts. *Teaching Mathematics and its Applications*, 25, 23–32.

Garfield, J. & Ben-Zvi, D. (2005). A Framework for Teaching and Assessing Reasoning about Variability. *Statistics Education Research Journal*, 4(1), 92–99. http://www.stat.auckland.ac.nz/serj

Garfield, J., delMas, R. & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM Mathematics Education*. DOI 10.1007/s11858-012-0447-5. Springer.

Henry, M. (2001). Notion de modèle et modélisation dans l'enseignement. In *Commission Inter-IREM Statistique et Probabilités, Autour de la modélisation en probabilités*. 149–159. Besançon, France: PUFC.

Jackson, J., Dukerich, L., & Hestenes, D. Modeling Instruction: an effective model for science education. (2008). *Science Educator*, 17(1), 10–17.

Jong, M. S., Shang, J., & Lee, F. (2010). Constructivist Learning Through Computer Gaming. In M. Syed (Ed.), *Technologies Shaping Instruction and Distance Education: New Studies and Utilizations* (pp. 207-222). Hershey, PA: Information Science Reference. doi:10.4018/978-1-60566-934-2.ch014

Lesh, R., Hoover, M., Hole, B., Kelly, A., & Post, T. (2000). Principles for developing thought-revealing activities for students and teachers. In A. E. Kelley & R. A. Lesh, eds., *Handbook of Research Design in Mathematics and Science Education* (591–646). Mahwah, NJ: Lawrence Erlbaum Associates.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the Evaluation of Systemic Science Education Reform: Searching for Instructional Sensitivity. *Journal of Research in Science Teaching*, 39(5), 369–393.

Shaughnessy, J. M. Ciancetta M., & Canada, D. (2004). Types of Student Reasoning on Sampling Tasks. *28th Conference of the International Group for the Psychology of Mathematics Education*, 4, 177–184.

Shaughnessy, J. M. (1992). Research in Probability and Statistics: Reflections and Directions. In D. A. Grouws (Ed.), *Handbook of Research on Teaching Mathematics and Learning*. New York: Macmillan.

Tversky, A. & Kahneman, D. (1974). Judgement under Uncertainty: Heuristics and Biases. *Science* 185, 1124–1131.