

# 10

## DOES PLAYING WORLD OF GOO FACILITATE LEARNING?

*Valerie J. Shute and Yoon Jeon Kim*

Failure is the opportunity to begin again, more intelligently.

(Henry Ford)

### Introduction

There has been a great deal of interest in video games and learning in recent years (e.g., Gee, 2003; Prensky, 2006; Shaffer, Squire, Halverson, & Gee, 2005; Shute, Rieber, & Van Eck, in press). A large part of this interest is motivated by frustration with the current education system and a desire for alternative ways of teaching and learning. Scholars in this games-learning arena argue that current schools in the United States do not adequately prepare kids for success in the 21st century. That is, learning in school is still heavily geared toward the acquisition of content, with instruction too often abstract and decontextualized, and thus not suitable for this age of complexity and interconnectedness (Shute, 2007).

In contrast to what children do inside of school, many of these same children spend countless hours playing fairly complex and challenging video games. In Ito and her colleague's 3-year ethnographic study (2010), they report that playing video games with friends and family is a large and normal part of the daily lives of youth. They also contend that playing video games is not solely for entertainment purposes. In fact, many youth enthusiastically participate in online discussion forums to share their knowledge and skills about a game with other players, or seek help in relation to challenges when needed. Kids use a variety of video- and picture-editing tools to share their playing on the Internet, and sometimes even learn how to modify the game (i.e., *modding*), which requires advanced computer technology skills.

**244** Valerie J. Shute and Yoon Jeon Kim

The main claim of researchers in the area of games and learning is that computer (or video) games can facilitate learning, because games provide a rich, interesting context, conducive for learning to occur (e.g., Gee, 2003; [[AQ61]]Shaffer et al., 2006). In addition to establishing context, well-designed games share many of the same features as exemplary learning environments. Some of these features include interactivity, immediate and ongoing feedback, adaptive levels of challenge, and complex problems with specific goals (Gee, 2003; Shute & Torres, in press). Well-designed games thus have the potential to elicit active and critical thinking and learning skills (e.g., Gee, 2003; Shute et al., in press; Shute, Ventura, Bauer, & Zapata-Rivera, 2009).

### **Learning Theory**

To support the claim that well-designed games are effective learning environments, we need to examine if and how people learn in those environments, and we need to be clear about what we mean by “learning.” The most widely accepted learning theory for games-and-learning research is *situated cognition* (e.g., Barab et al., 2007; Gee, 2003; Lave & Wenger, 1991; Shaffer et al., 2005). Situated cognition defines human learning, thinking, and problem solving as being embodied within a context. People learn through active experiences and critical interpretation of their experiences via personal reflection and interpersonal discussion.

In addition, *people learn in action* in video games (Gee, 2008b; 2010; Salen & Zimmerman, 2005). That is, people interact with all aspects of the game and take intentional actions within the game. For its part, the game continuously responds to each action, and, through this process, the player gradually creates meaning. Clearly, how people are believed to learn within video games contrasts to how people typically learn at school, which often entails memorization of decontextualized and abstract concepts and procedures (Shute et al., 2009).

### **Purpose and Organization of the Chapter**

The main purpose of this chapter is to provide an example of an evidence-based assessment used within a commercial game to examine any learning of educationally valuable knowledge and skills that may take place during gameplay. Our beliefs motivating this research are twofold: (a) it is important to develop valid models and assessments for complex knowledge and skills that are required for success in the 21st century; and (b) assessments can be embedded within video games to support such skills that are currently not being assessed and supported. Our goal is to illustrate how people can develop educationally valuable skills (e.g., problem solving and causal reasoning) by playing a well-designed video game that is not explicitly developed for educational purposes.

The organization of this chapter is as follows. First, we briefly summarize a few examples of games and learning research to provide a feel for what's currently being done in the area. Second, we review assessment research conducted in relation to games. This is followed by an overview of a particularly effective assessment approach called evidence-centered design (ECD). The bulk of the chapter describes the game that we used to illustrate our claims—*World of Goo* (2008)—focusing our attention on how we developed and applied our ECD models to the game. We touch on our use of Bayesian networks (Pearl, 1988; Pearl & Russell, 2003) to tie the models together. And finally, we discuss findings from an exploratory assessment study, and close with implications for future research.

## Background

### *Examples of Games and Learning*

Our first example illustrates how kids learn science content and inquiry skills as situated in an online game called *Quest Atlantis: Taiga Park* (Barab et al., 2007; Shute et al., 2010). Players enter the game as an assistant to the park ranger. As such, they have to explore Taiga Park to investigate and understand why the fish are dying in the river. Players are immediately and actively engaged in various tasks, such as collecting water samples, interviewing stakeholders, creating hypotheses, and solving problems—large and small. They also travel across time via a time machine to see the consequences of their decisions and actions. From this gameplay, the players experience (via data collection and hypothesis making/revising) scientific inquiry and evolve in their understanding about how certain science concepts are related to each other (e.g., sediment in the water from the loggers' activities causes increased water temperature, which causes decreased dissolved oxygen, which causes the fish to die). Most importantly, players have many opportunities to reflect on their actions and make meaning from their experiences.

Our next example relates to research conducted by Squire (2004), who employed a popular commercial game called *Civilization* in a world-history class for high-school students. He reported that students playing the game developed deep and complex understanding and language in relation to world history.

Not only can playing video games facilitate learning of academic subjects, it also can facilitate the acquisition of complex thinking skills (e.g., problem solving and systems thinking). For instance, as Torres (2009) reported, students who played *Gamestar Mechanic* (i.e., a video game where kids play the role of game designer) developed systems thinking skills, and students playing Taiga Park similarly acquired systems thinking skills (Shute et al., 2010).

246 Valerie J. Shute and Yoon Jeon Kim

### **Assessment Research**

Despite these preliminary, promising results, Gee (in press) and many others (e.g., Cannon-Bowers, 2006; van Eck, 2007) are quick to point out that there is still limited empirical evidence to support the range of learning-from-games claims. In an effort to begin to validate some of the claims and to provide evidence for why and how video games are good for learning, some researchers are paying greater attention to assessment (e.g., Rupp, Gushta, Mislevy, & Shaffer, 2010; Shaffer, 2006; Shute, in press).

Two examples highlight the importance of assessment research as applied to video games. First, Shute and colleagues (2009) describe *stealth assessment*—assessment that is seamlessly woven into a learning environment, such as games or simulations. They illustrate the approach within a commercial video game called *Oblivion*. Their example demonstrates how assessment can be situated within a game environment and be used to provide formative feedback and thus support learning of important competencies that are essential for success in the 21st century (e.g., creative problem solving).

Second, Shaffer and his colleagues (Shaffer et al., 2009) describe another approach to assessment in video games called *epistemic network analysis*. This is intended to be used within epistemic games such as *Urban Science* and *science.net* and represents another assessment approach for use in video games. Epistemic games are designed to allow players to think and act like domain-specific experts, such as urban planners and science journalists. Their example provides a way to understand players' growth in relation to the skills, knowledge, identities, and values of experts by playing epistemic games.

One common denominator underlying the two examples described above (i.e., Shute et al., 2009; Shaffer et al., 2009) is the use of an assessment design framework called ECD (Mislevy, Steinberg, & Almond, 2003).

### **Overview of Evidence-Centered Design**

ECD is a conceptual framework that can be used to develop assessment *models*, which in turn support the design of valid assessments. The goal is to help assessment designers coherently align (a) the claims that they want to make about learners, and (b) the things that learners say or do in relation to the contexts and tasks of interest (for an overview, see Mislevy & Haertel, 2006; Mislevy et al., 2003). There are three main theoretical models in the ECD framework: competency, evidence, and task models.

The competency model consists of student-related variables (e.g., knowledge, skills, and other attributes) about which we want to make claims. For example, suppose that you wanted to make claims about a student's ability to "design excellent presentation slides" using MS PowerPoint. Your competency-model variables (or nodes) would include technical as well as visual design skills. Your

evidence model would show how, and to what degree, specific observations and artifacts can be used as evidence to inform inferences about the levels or states of competency–model variables. For instance, if you observed that a learner demonstrated a high level of technical skill but a low level of visual design skill, you may estimate her overall ability to design excellent slides to be approximately “medium”—if both the technical and aesthetic skills were weighted equally. The task model in the ECD framework specifies the activities or conditions under which data are collected. In our current PowerPoint example, the task model would define the actions and products (and their associated indicators) that the student would generate, comprising evidence for the various competencies.

There are two main reasons why we believe that the ECD framework fits well with the assessment of learning in video games. First, in video games, people learn in action (Gee, 2003; Salen & Zimmerman, 2005). That is, learning involves continuous interactions between the learner and the game, so that learning is inherently situated in context. Therefore, the interpretation of knowledge and skills as the products of learning cannot be isolated from the context, and neither should assessment. The ECD framework helps us to link what we want to assess and what learners do in complex contexts. Consequently, an assessment can be clearly tied to learners’ actions within video games, and can operate without interrupting what learners are doing or thinking (Shute, in press).

The second reason that ECD is believed to work well with video games is because the ECD framework is based on the assumption that assessment is, at its core, an evidentiary argument. Its strength resides in the development of performance-based assessments where what is being assessed is latent or not apparent (Rupp et al., 2010). In many cases, it is not clear what people learn in video games. However, in ECD, assessment begins by figuring out just what we want to assess (i.e., the claims we want to make about learners), clarifying the intended goals and outcomes of learning.

We will now see how this type of assessment approach can be applied within an existing game to determine what, if anything, of value is learned during gameplay.

### ***World of Goo***

*World of Goo* is a physics-based puzzle game where players utilize various types of “goo ball” to build different structures to reach suction pipes (Davidson, 2009). Each level consists of a different environment as well as the required number of balls to complete the level.

Depending on the nature of the environment, there are many forces working against each other, such as gravity and buoyancy. Their combination determines the stability (i.e., equilibrium) of the goo-ball structures. Thus, a player in the *World of Goo* needs to effectively solve a number of complex and novel problems.

248 Valerie J. Shute and Yoon Jeon Kim

One level of the game that we showcase in this chapter is called “Fisty’s Bog” (for more on this level, see [http://worldofgoo.wikia.com/wiki/Fisty's\\_Bog](http://worldofgoo.wikia.com/wiki/Fisty's_Bog)).

Similar to other levels in the game, players need to build a structure (in this case, a bridge) by connecting goo balls together and strategically affixing balloons to the structure. Fisty’s Bog begins with a player figuring out what she is supposed to do to win the level. There are no explicit hints given to the player about what she is supposed to do. The only way the player can figure out the goal of the level is by engaging in exploratory behaviors, observations, reflections, and continuous hypothesis making and testing.

### ***ECD Models Applied***

Assessment in Fisty’s Bog can support diagnosis of player performance and also provide the basis for formative feedback to the player—either explicitly (to the player or teacher) or upon demand. That is, specifying relevant knowledge and skills within the competency model at a sufficiently fine grain size for Fisty’s Bog allows for inferential judgments to be made related to a learner’s performance.

Consider a player who did not win the level on her first attempt. Maybe the player simply did not understand the goal, or perhaps she understood the goal, but experienced procedural difficulties when trying to execute a solution. That level of detail (i.e., conceptual misunderstanding vs. procedural problems) can be much more instructionally helpful than just informing her that she failed the level.

The assessment design for Fisty’s Bog began with a cognitive task analysis (see Mislevy, Steinberg, Breyer, Almond, & Johnson, 1999) consisting of four parts. First, we met with a small group of *World of Goo* experts ( $n = 3$ ) gathered to discuss relevant knowledge and skills needed to succeed in Fisty’s Bog. Second, think-aloud protocols were collected from five individuals with varying levels of expertise in Fisty’s Bog, from complete novice to expert. The main goal of this phase was to observe how people with different levels of expertise play Fisty’s Bog. Third, the think-aloud protocols from the five players were analyzed, and we extracted features of performance that differentiated levels of proficiency. Finally, the collected information from the analysis served as the basis for coherently structuring the ECD models, described next.

### ***The Competency Model of Fisty’s Bog***

Results from our cognitive task analysis yielded three main competency-model variables that are required for successful understanding and gameplay in Fisty’s Bog: problem-solving skill, causal reasoning skill, and knowledge of static equilibrium. Each is described in turn.

- *Problem solving*: Problem solving consists of figuring out what the problem is, as well as the goal, and then coming up with appropriate solutions. Novice

Does Playing *World of Goo* Facilitate Learning 249

players of *Fisty's Bog* were not able to immediately identify what they were expected to do in *Fisty's Bog*. Furthermore, novice players were unable to identify all of the available resources and obstacles relating to the problem. In contrast, expert players quickly figured out the problem, identified helpful resources and obstacles to surmount, and were able to hypothesize solutions to solve the problem.

- *Causal reasoning*: From our review of the think-aloud protocols collected during the cognitive task analysis, we noticed that novice players were unable to identify all the variables that were interrelated in *Fisty's Bog*. Moreover, their explanations of the relationships tended to be rather simplistic and incomplete. Conversely, experts identified most of the variables that were interrelated, and their explanations for the causal relationships were fairly complex and comprehensive.
- *Knowledge of static equilibrium*: The physics principle applied in *Fisty's Bog* is static equilibrium. We hypothesized that a player who understood the conditions of static equilibrium would be able to readily attain balance of the bridge-like structure by manipulating the source of forces. Initially, it was not clear whether or not knowledge of static equilibrium is a critical competency in the solution of the level, to the same degree as problem solving and causal reasoning. That is, some players who had no formal knowledge of static equilibrium were still able to win the level, if they showed sufficient problem-solving and causal-reasoning skills. However, further analysis revealed that, even though the players didn't use formal physics terms, they still had a reasonable conceptual or intuitive understanding of static equilibrium that was expressed via informal language. For example, a player said, when he was asked what he was doing, "I am trying to keep it balanced . . . trying to build sort of a zigzag structure. Once it [the bridge] gets too long, it gets heavy and falls . . . need to add balloons to balance." We did, however, observe that a player who was an expert in physics more strategically distributed goo balls and balloons, compared with other players without the formal knowledge of static equilibrium.

The competency model for *Fisty's Bog* is shown in Figure 10.1. To win *Fisty's Bog*, the first thing a player needs to do is analyze the given problem. That is, a player should be able to state the mission or goal of the particular level in the game. Subsequently, a player should be able to indicate all available resources, including goo balls, balloons, the sign, and even the subtitle of the game (which, in this case was, "Not too high, not too low"). In addition, one should be able to indicate that upper and bottom spikes serve as serious obstacles as they can pop the inflated balloons (upper spikes) or kill goo balls (bottom spikes). Finally, one needs to hypothesize potential solutions. This solution-generation process is iterative, because a player needs to carefully monitor how the system reacts when the hypothesized solution is applied and make adjustments accordingly.

250 Valerie J. Shute and Yoon Jeon Kim

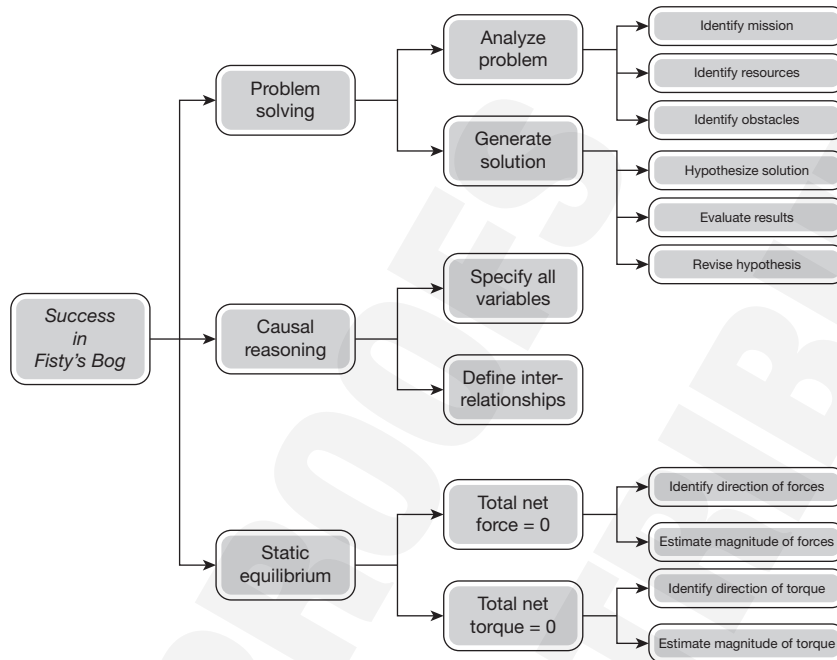


FIGURE 10.1 Competency model of Fisty's Bog.

Players should also be able to identify how the elements of the game and their interrelationships affect gameplay. For example, a player needs to understand that there are certain causal relationships among goo balls, such as sagging of the goo-ball structure, locations of balloons to counteract sagging, and the overall balance of the structure. In addition, having some informal or intuitive understanding of static equilibrium helps players to evenly distribute the weight in the structure, which helps players to build a stable structure.

### The Evidence Model of Fisty's Bog

The evidence model in ECD determines how the observed actions in the game can be used as evidence to update the current states of the competency-model variables. That is, the evidence model statistically links specific observations with associated variables of the competency model, which are unobservable. The evidence model is composed of three parts: (a) *scoring rules* for indicators (i.e., how to score what a player does or says in relation to playing the game), (b) *scores* (e.g., values assigned to players' answers to questions and their causal maps), and (c) *statistical models* (e.g., probability distributions for competency-model variables for data accumulation). For example, when a player accurately identified the mission of Fisty's Bog, she would obtain a score of 1 (with a range of 0–1).



Her score is then fed back to the corresponding competency-model variable (i.e., identify the mission) using a statistical accumulation process (e.g., Bayes net).

### *The Action Model of Fisty's Bog*

Shute et al. (2009) renamed the task model of the ECD framework to “action” model when used in games, because an action in a game is what a learner does by interacting with the environment to solve problems. The action model defines the sequence of actions that a player takes during gameplay, and each action’s indicators of success (and failure). Some actions are required to be sequential to proceed within the mission, while some actions can repeatedly occur. For example, in Fisty’s Bog, the first required action is to identify the goal of the current level or quest. Next, a player needs to repeatedly array goo balls and balloons to build a bridge-like structure. Table 10.1 illustrates some representative actions that a player needs to take, along with their associated indicators to succeed in Fisty’s Bog.

The ECD-based models need to be tightly aligned so that they can be used for reasoning about players’ proficiency in Fisty’s Bog. To join all the models together, we employed a Bayesian network (or Bayes net).

**TABLE 10.1** Action and indicators

<i>Action</i>	<i>Indicators</i>
Identify mission	State that the goal of this level is to reach up to the suction pipe by creating a structure that reaches to the suction pipe, and at least 6 goo balls need to be sucked into the pipe
Identify available resources	Identify the use of balloons Identify the use of goo balls to construct a bridge Identify the use of reset button (i.e., flying flies) Identify the purpose of the suction pipe Understand the meaning of the hint (not too high, not too low) Identify the use of the sign (additional hints)
Identify obstacles	Avoid upper spikes Avoid bottom spikes
Hypothesize solutions	State hypothesized solutions
Evaluate results	Analyze and articulate reasons for success Analyze and articulate reasons for failure
Revise hypothesis	Modify hypothesized solution to solve the problem in subsequent gameplay

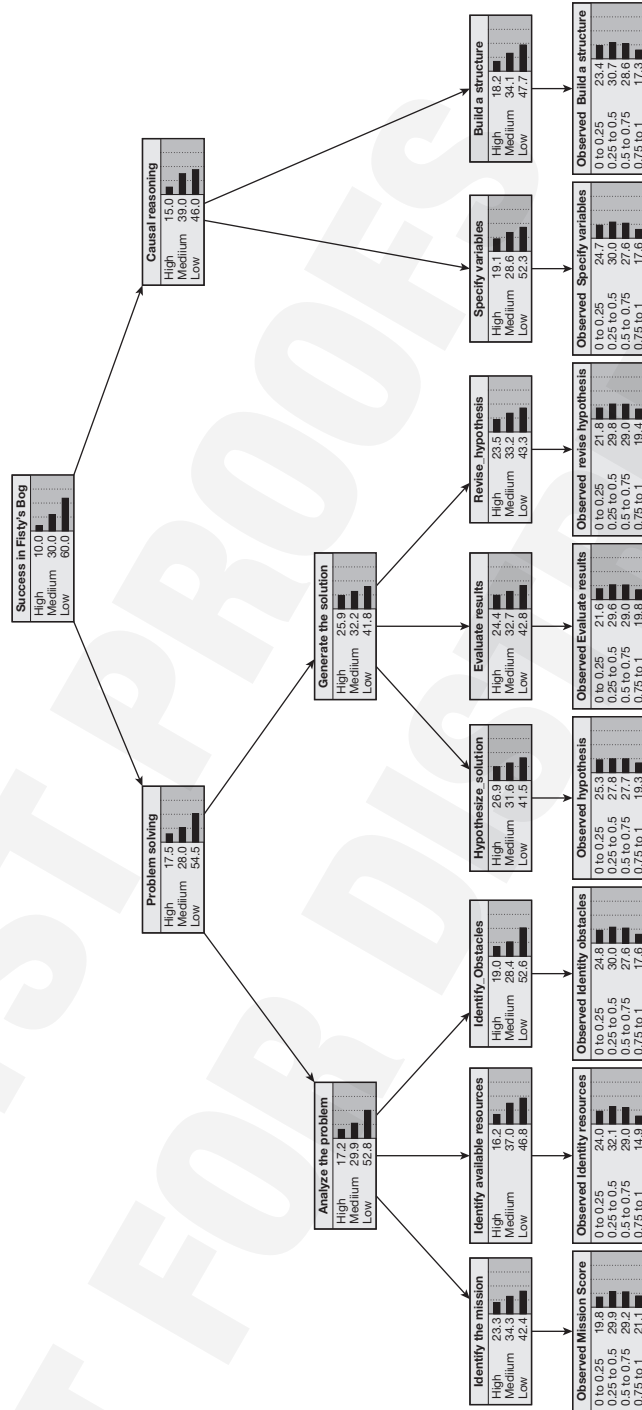


FIGURE 10.2 Bayesian model used to instantiate our ECD-based conceptual framework.

### **Success in *Fisty's Bog***

Similar to the examples described by Shute and colleagues (e.g., Shute et al., 2009, 2010), we formalized our competency and evidence models by employing a Bayes net. In this study, only two competency-model variables (e.g., problem solving and causal reasoning) were formalized and expanded. Modeling the third variable—static equilibrium—was not done because of time and computational constraints. That is, because this was an exploratory study, we first wanted to test parts of the model (problem solving and causal reasoning) that we believed exerted more influence on one's success in the game. Also, as mentioned, almost all of the individuals we observed during our cognitive task analysis phase seemed to have an intuitive understanding of static equilibrium coming into the game.

A Bayes net links together actions, evidence, and claims about competencies, which allows for probability-based reasoning of the learner's performance (Mislevy, 1994). A Bayes net can integrate evidence from a learner's performance and produce marginal probabilities for each variable of the competency model. The a priori probabilities (i.e., the priors) for the conditional probability tables of the Bayes net were obtained from the cognitive task analysis. Figure 10.2 shows our initial Bayesian model instantiating our ECD-based conceptual framework.

In the following section, we describe how our ECD-based assessment was administered and then interpret results from an exploratory study using *World of Goo*.

## **Exploratory Study**

### **Method**

Three undergraduate students (ages 18–22) and one graduate student (age 38) participated in this exploratory study. None had prior experience playing *World of Goo* (see Appendix A for the survey questions used to recruit participants for both the cognitive task analysis and the exploratory study). Before the participants were individually tested, we provided a warm-up exercise for the think-aloud procedure. Once each participant started playing the game, interactions with the game were automatically recorded by a screen-capturing application called Fraps (2010). This allows synchronizing of the verbal protocol with players' actions. Some structured questions were also asked while they were playing the game (Appendix B), and a behavior checklist was utilized (Appendix C) to obtain additional information regarding specific indicators.

Once each player completed or elected to quit the game, he or she was asked to draw a map explaining the causal relationships among all of the variables in the game (Appendix D). We then transcribed and analyzed the verbal protocols of the players. Table 10.2 summarizes the descriptive data of the four participants.

254 Valerie J. Shute and Yoon Jeon Kim

**TABLE 10.2** Descriptive data about the participants

	<i>S1</i>	<i>S2</i>	<i>S3</i>	<i>S4</i>
Age	20	18	22	38
Gender	M	F	M	M
How often do you play video games?	Almost every day	Once a month	Never	Never
How good you are at playing games in general?	Very good	Fair	Good	Poor
Number of attempts	2	4	3	3
Completed the level	Yes	No	No	Yes

### Results

To demonstrate how our ECD models were used to provide information about one's proficiency in *Fisty's Bog*, we compared two participants relative to their performance in the game (i.e., *S3* and *S4*). Figures 10.3 and 10.4 indicate the overall estimated proficiency levels of *S3* and *S4* when the evidence is integrated into the Bayes net.

As indicated in Table 10.2, *S3* didn't win the level after three attempts, while *S4* won the level on his third attempt. Based solely on this information, one might conclude that *S4* is a better player than *S3* because he won the level, and that *S4* is more likely to have higher problem-solving and causal-reasoning abilities than *S3*. However, when the evidence model was updated, we were able to make more valid assessments of their respective competencies, and here is why.

### Problem Solving

Both *S3* and *S4* scored perfectly (with a value = 1.0) relative to being able to identify the mission of the game. That is, they stated that the goal of the game is to build a bridge-like structure using goo balls and pink balloons to reach the suction pipe at the other end. However, when it came to identifying resources and obstacles, *S4* did not identify certain resources, such as the hint from the subtitle of the game and the utility of the reset buttons. He also didn't identify the bottom spikes as an obstacle. In contrast, *S3* identified all six resources and indicated both upper and bottom spikes as obstacles. Therefore, the overall posterior probability of *S3*'s "analyze the problem" competency was estimated as high ( $p = .93$ ), whereas *S4*'s posterior probability of being high in relation to that competency was less,  $p = .63$ .

So, both *S3* and *S4* played *Fisty's Bog* three times, and *S4* won the level, whereas *S3* did not. And we found that *S3* identified more resources and obstacles

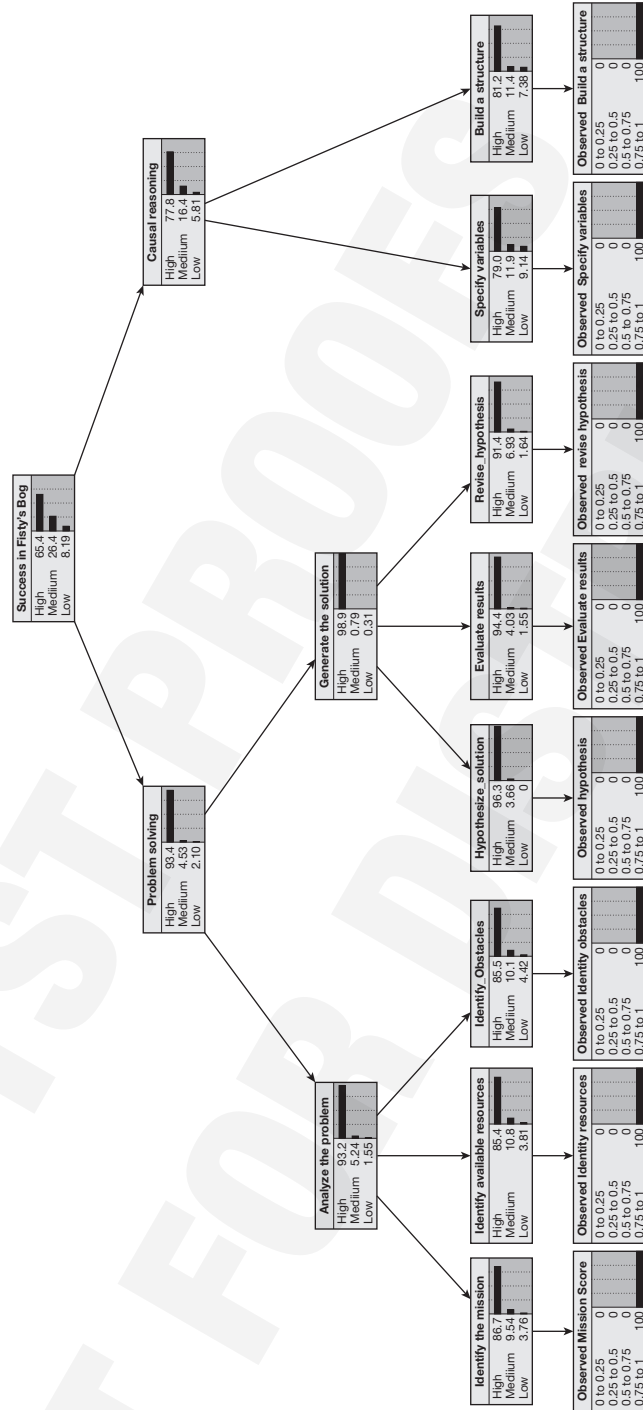


FIGURE 10.3 S3's posterior probabilities after updating the evidence model.

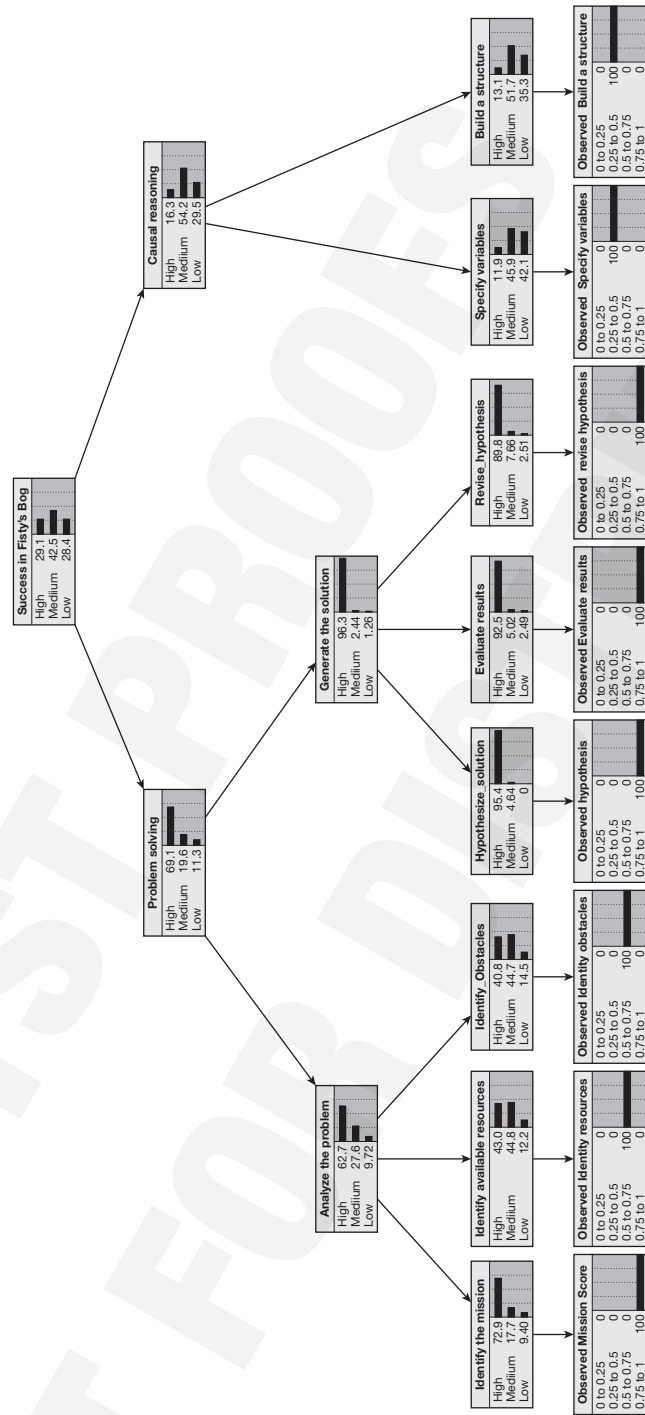


FIGURE 10.4 S4's posterior probabilities after updating the evidence model.

Does Playing *World of Goo* Facilitate Learning 257

compared with S4. Does that finding have any bearing on learning, especially in terms of anything noteworthy? To answer this question, we further analyzed the data obtained from the captured videos, focusing on how much time each player spent in the game during their first attempt. The reason that we looked at their first-attempt data is because that is when players are more likely to be engaged in exploratory behaviors (e.g., random clicking around until something changes) compared with more purposeful actions. We found that the total amount of time S3 spent during his first attempt was 11 minutes, 51 seconds, whereas S4 spent *less than half* the time of S3, at 5 minutes, 38 seconds. This suggests that S3 spent more time exploring the game environment and elements compared with S4.

We further analyzed the verbal protocols of the two players during their first attempt. We wanted to see if there were any qualitative differences in terms of how they were respectively engaged. Consistent with the amount of the time they spent, S3 paid attention to every single element in the game. For example, he read and reread the hint sign three times, whereas S4 only clicked it once. S3 also tried to make sense out of the sign that was intentionally vague, whereas S4 simply said it didn't give any tips. S3 carefully evaluated each action he made, saying things such as, "I think maybe I did something right," and "Something just happened and I don't know why yet." In contrast, S4 clicked around till something accidentally happened. And, instead of trying to explain why something was changed in the game, he repeated his apparently aimless clicking. Even though these differences between the two players did not predict who eventually won the level, it helped shed light on who was more actively and critically engaged during gameplay (i.e., S3).

### *Causal Reasoning*

The causal-reasoning estimate was based on the completeness and accuracy of the causal maps that S3 and S4 produced after playing the game. Causal maps represent the players' causal knowledge about the game components. The quality of the produced causal maps is determined by the number of relevant variables included, the number of links between variables, and appropriate directions of the links (Spector, Christensen, Sioutine, & McCormack, 2001). For comparison purposes, we developed and used an *expert* (or reference) causal map that was produced from the results of the cognitive task analysis (see Figure 10.5).

When we examined the causal maps from both players relative to the expert map, we found that S3 identified 10 relevant variables that overlapped with the expert's map, whereas S4 only listed 3 shared variables. In addition, S3's map was more complex than S4's, as he linked all the variables with appropriate directions. Consequently, S3's diagram was seen as more complex and complete than S4's. When those data were incorporated into the Bayes net (see Figures 10.3 and 10.4), we see that, indeed, S3 has a higher estimated causal-reasoning skill than S4

258 Valerie J. Shute and Yoon Jeon Kim

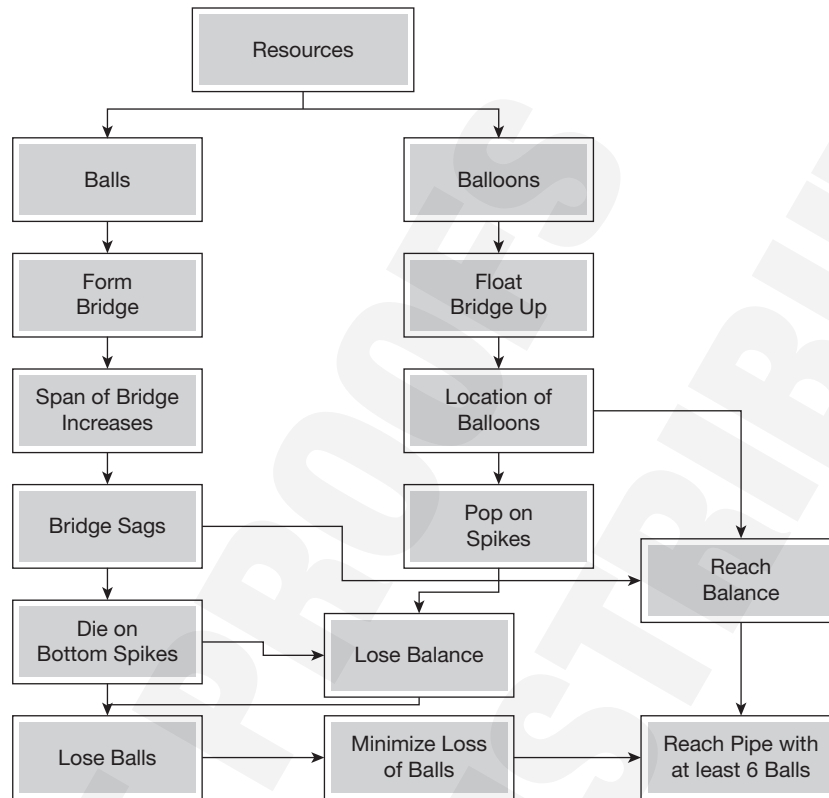


FIGURE 10.5 Expert map.

( $p = .78$  versus  $p = .16$ , respectively, in terms of the probability of being “high” on this variable).

S3’s map indicates his understanding of *Fisty’s Bog* as a system where various components of the game have causal relations. Furthermore, even though he did not explicitly use physics terms to explain these relationships, his map implies that he was able to induce the physics principle of static equilibrium (i.e., balancing two forces to make the applied net force equal to 0). On the other hand, S4’s map is comparatively simplistic, suggesting that he failed to fully understand or explicate the various relations among game elements.

As illustrated by the comparisons of S3 and S4, just completing or winning *Fisty’s Bog* does not automatically ensure the development of causal-reasoning and/or problem-solving skills. Rather, it depends on how actively and critically a learner (a) was engaged in the task (i.e., playing the game), and (b) reflected on the successes and failures during the game-play process. That is, even though S3 did not win the level, he was able to develop a higher estimated level of problem solving and causal reasoning relative to S4.



Does Playing *World of Goo* Facilitate Learning 259

How can this result be explained? There are two important factors that affect meaningful learning in video games: reflection and failing (Gee, 2005). As Gee emphasized, the *reflection* on actions (and reactions) that take place during gameplay is crucially important for learning in video games. Furthermore, in well-designed games, *failing* is a critical part of learning. Players learn to revise their actions only as a result of failing. Therefore, “learning” in video games cannot be judged solely on the basis of whether one masters the game or wins a level. Evidently, S3 and S4 differed quite a bit in terms of their respective reflections of failure and success. This can be clearly seen in their think-aloud protocols following various attempts to win the level. When S3 failed in his first attempt, he analyzed his failure as follows:

S3: Initially, I had no idea how to approach this game, but later I realized that you need to connect the goo balls together and use the balloons to keep the balls floated off the spikes . . . you need to make the thinnest bridge as you can across with most balloons spread out . . . so it doesn't droop. But I am out of balls now . . . I let my balloons pop and I'm kind of stuck in the middle like half-way.

On the other hand, when S4 was asked why he failed to win the level at his first attempt, he simply responded, “Because I didn't collect anything. I'm supposed to collect something.”

Comparing those two responses, it seems that S3 understood the problem that he needed to solve and derived a viable solution better than S4 did after the first attempt. Moreover, S3 knew what he should do on his second attempt to win the level. Once both of them failed their second attempts, they were again asked why, to which they replied:

S3: . . . I'm stuck because I'm out of little white flies [reset buttons] to help me get another balloon or goo balls. This time, I tried to balance out adding goo balls and balloons at the same time . . . it either rises up too fast and popped balloons, or sinks down too fast and kills goo balls . . . [I] didn't have enough balloons or goo balls to continue.

S4: It took too long to realize the goal of the game, and my strategy wasn't good enough. If I play again, I know that I need to be careful with using the goo balls, so they don't run out.

As indicated from their answers, S3 appears to have a somewhat deeper understanding of the game than S4, and this explains why he was able to score higher than S4 in terms of problem-solving and causal-reasoning estimates, even though he didn't win the level.

260 Valerie J. Shute and Yoon Jeon Kim

## Discussion

By implementing an ECD-based assessment linked to an existing causal game, we were able to examine the acquisition of certain knowledge and skills during gameplay. One interesting finding was that simply playing the game (and even winning the level) does not automatically translate into learning valuable knowledge and skills, especially to the degree, or of the kind, that can be transferred outside of the game setting. This was illustrated by the assessment results of S3 and S4. That is, even though video games have great potential as a rich context for learning, the success of learning ultimately depends on *how* one interacts with the game. Without such an understanding about the features of learners and video games, one cannot support the claim that playing video games facilitates learning.

We believe that our exploratory study has two implications to the field of video games and learning. First, employing this type of evidence-based assessment can help us to better understand the underlying learning theory in video games. That is, the present study highlights the importance of *failure* in relation to the learning processes in video games. Rather than construing it as a negative thing to be avoided (as in traditional education settings), failure should be viewed as a catalyst for reflection (e.g., *Why didn't this work, and what can I do differently next time?*). Second, ECD-based assessment can help us to make valid claims about what people do learn by playing video games, especially when the games are not explicitly educational.

We illustrated why assessment is so important in relation to video games. That is, the obtained assessment information can be used to support the development of a wide range of educationally valuable skills, including those that are not typically assessed and supported in current educational systems (e.g., causal reasoning). Because competency models are always up to date regarding a player's learning estimates, the information may ultimately be used diagnostically to direct targeted interventions or provide specific formative feedback ([AQ62]Shute, 2008). Moreover, information from the assessment can be used to help a player to engage in deep and critical reflection of one's actions and their consequences. However, even though reflection appears to be a key part of successful learning within video games, most game designs do not explicitly include features that facilitate reflection.

In conclusion, we described an assessment that was developed based on the ECD framework, and we illustrated how it could be used to provide the basis for valid inferences about players' learning in a particular, commercial video game. Although our preliminary results are encouraging, there is still much work to be done. More empirical studies need to be conducted to ascertain what comprises "good video games," and how they may be harnessed to facilitate people's active and critical learning and thinking (Gee, in press). We also recommend more effort by the research community to focus on identifying, modeling, and assessing what

people learn. Ultimately, this can lead to a repository of valid competency models that may be reapplied in various learning environments—including other games. Finally, more examples of assessment in video games need to be tested to determine which aspects of those models are suitable, or not, for video games, or suitable for particular genres of games.

In closing, we note that the old expression, “It doesn’t matter if you win or lose—it’s *how* you play the game” takes on a whole new meaning relative to learning and the findings in this chapter. In particular, losing, in a curious way, may actually be a form of winning when it comes to video games and learning.

## References

- Barab, S. A., Zuiker, S., Warren, S., Hickey, D., Ingram-Goble, A., Kwon, E.-J., Kouper, I., & Herring, S. C. (2007). Situationally embodied curriculum: Relating formalisms and contexts. *Science Education*, 91(5), 750–782.
- Cannon-Bowers, J. (2006, March). *The state of gaming and simulation*. Paper presented at the Training 2006 Conference and Expo, Orlando, FL.
- Davidson, D. (2009). From experiment gameplay to the wonderful world of goo and how physics is your friend. In D. Davidson (Ed.), *Well played 1.0: video games, values and meaning* (pp. 334–367). Pittsburgh, PA: ETS Press.
- Fraps [Computer Software] (2010). (Version 3.2.2.) Beepa.
- Gee, J. P. (2003). What video games have to teach us about learning and literacy? New York, NY: Palgrave/Macmillian.
- Gee, J. P. (2005). Learning by design: good video games as learning machines. *E-Learning and Digital Media*, 2(1), 5–16 Retrieved from <http://dx.doi.org/10.2304/elea.2005.2.1.5>.
- [[AQ63]]Gee, J. P. (2008a). Learning and games. In S. Katie (Ed.), *The ecology of games: connecting youth, games, and learning* (pp. 21–40). The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. Cambridge, MA: The MIT Press.
- Gee, J. P. (2008b). Video games and embodiment. *Games and Culture*, 3(3–4), 253–263.
- Gee, J. P. (2010). Human action and social groups as the natural home of assessment: Thoughts on 21st century learning and assessment. In V. J. Shute & B. J. Becker (Eds.), *Innovative assessment for the 21st century: Supporting educational needs*. New York, NY: Springer-Verlag.
- Gee, J. P. (in press). Reflections on empirical evidence on game and learning. To appear in S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction*. Charlotte, NC: Information Age Publishers.
- Ito, M., Baumer, S., Bittanti, M., Boyd, D., Cody, R., Herr-Stephenson, B., et al. (2010). *Hanging out, messing around and geeking out: kids living and learning with new media*. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. Cambridge, MA: The MIT Press.
- [[AQ63]]Jenkins, H. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. Cambridge, MA: The MIT Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

**262** Valerie J. Shute and Yoon Jeon Kim

- Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, *59*, 439–483.
- Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2004). *A brief introduction to Evidence-Centered Design* (CSE Report 632). CA: Center for Research on Evaluation, Standards, and Student Testing (ED483399).
- Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessment. *Measurement: Interdisciplinary research and perspective*, *1*(1), 3–62.
- Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (1999). A cognitive task analysis with implications for designing simulation-based performance assessment. *Computers in Human Behavior*, *15*, 335–374.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo, CA: Morgan Kaufmann.
- Pearl, J., & Russell, S. (2003). Bayesian networks. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed., pp. 157–160). Cambridge, MA: MIT Press.
- Prensky, M. (2006). *Don't bother me mom, I'm learning!: How computer and video games are preparing your kids for 21st century success and how you can help!* St. Paul, MN: Paragon House.
- Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *Journal of Technology, Learning, and Assessment*, *8*(4). Retrieved from <http://escholarship.bc.edu/jtla/vol8/4>.
- Salen, K., & Zimmerman, E. (2005). Game design and meaningful play. In J. Raessens & J. Goldstein (Eds.), *Handbook of computer game studies* (pp. 59–80). Cambridge, MA: MIT Press.
- Shaffer, D. W. (2006). *How computer games help children learn*. New York, NY: Palgrave Macmillan.
- Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Franke, K., Rupp, A. A., & Mislevy, R. J. (2009). Epistemic Network Analysis: A prototype for 21st Century assessment of learning. *The International Journal of Learning and Media*, *1*(2), 33–53.
- Shaffer, D. W., Squire, K. A., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan*, *87*(2), 104–111.
- Shute, V. J. (2007). Tensions, trends, tools, and technologies: Time for an educational sea change. In C. A. Dwyer (Ed.), *The future of assessment: shaping teaching and learning* (pp. 139–187). New York, NY: Lawrence Erlbaum Associates, Taylor & Francis Group.
- Shute, V. J. (in press). Stealth assessment in computer-based games to support learning. To appear in S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction*. Charlotte, NC: Information Age Publishers.
- Shute, V. J., Rieber, L., & Van Eck, R. (in press). Games . . . and . . . learning. To appear in R. Reiser & J. Dempsey (Eds.), *Trends and issues in instructional design and technology* (3rd ed.). Upper Saddle River, NJ: Pearson Education, Inc.
- Shute, V. J. & Torres, R. (in press). Where streams converge: Using evidence-centered design to assess Quest to Learn. To appear in D. Robinson, J. Clarke-Midura, & M. Mayrath (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research*. New York, NY: Springer-Verlag.

Does Playing *World of Goo* Facilitate Learning **263**

- Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. J. Cody, & P. Vorderer (Eds.), *The social science of serious games: theories and applications* (pp. 295–321). Philadelphia, PA: Routledge/LEA.
- Spector, J. M., Christensen, D. L., Sioutine, A. V., & McCormack, D. M. (2001). Models and simulations for learning in complex domains: using causal loop diagram for assessment and evaluation. *Computers in Human Behavior*, *17*, 517–545.
- Squire, K. (2004). *Replaying history: Learning world history through playing Civilization III*. Indiana University: ProQuest Dissertations.
- [[AQ63]]Tobias, S., & Fletcher, J. D. (2007). What research has to say about designing computer games for learning. *Educational Technology*, *47*(5), 20–29.
- Torres, R. J. (2009). *Learning on a 21st century platform: Gamestar Mechanic as a means to game design and systems-thinking skills within a nodal ecology*. New York University: ProQuest Dissertations.
- Van Eck, R. (2007). Six ideas in search of a discipline. In B. Shelton & D. Wiley (Eds.), *The educational design and use of computer simulation games*. Boston, MA: Sense.
- World of Goo* [Video Game] (2008). 2D Boy.

264 Valerie J. Shute and Yoon Jeon Kim

## APPENDIX A

Your name: \_\_\_\_\_  
Gender: \_\_\_\_\_  
Age: \_\_\_\_\_  
Highest degree earned: \_\_\_\_\_  
Degree in: \_\_\_\_\_

**How often do you play video games? (Please tick one option)**

- Almost every day
- 3–4 times a week
- Once a week
- Once a month
- Never

**If you regularly play video games, what kinds of video games do you play? (Please tick ALL that apply)**

- Action (e.g., Street Fighter)
- Platform (e.g., Super Mario Bros)
- Role-playing (e.g., Dungeons & Dragons)
- Shooter (e.g., Halo)
- Simulation (e.g., The Sims)
- Strategy (e.g., Civilization)
- Others

**Can you name some of the games that you played in the past?**

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

**How good you think you are at playing video (computer) games in general?**

- Very good
- Good
- Fair
- Poor
- Very poor

## APPENDIX B

### Protocol Analysis prompts

In this experiment I am interested in what you say to yourself as you play the game that I will give to you. In order to do this I will ask you to THINK ALOUD as you play the game. What I mean by think aloud is that I want you to say out loud everything that you say to yourself silently. Just act as if you are alone in this room speaking to yourself. If you are silent for any length of time I will remind you to keep talking. You are also not allowed to ask me questions while you are playing the game. Do you understand what I want you to do?

Before we turn to the real experiment, I will give you a practice problem. I want you to talk while you do this problem. First I will ask you to multiply two numbers in your head.

So “think aloud” while you multiply 24 times 34.

Good! I think you are ready now. Let’s go ahead and start.

### Reminder: Keep talking

- 1 What is the goal of this level? (Identify mission)
- 2 How you think you can win this level? (Hypothesize solutions)
- 3 (once they succeed or fail) Can you explain why you could/could not win this level?
- 4 (once they succeed or fail) If you were given another chance, how would you play this time?
- 5 (At the end) Would you like to try again?





## APPENDIX D

Your name: \_\_\_\_\_

List all the variables (elements) that you identified to win Fisty's Bog

---

---

---

---

---

---

---

---

---

---

---

---

Draw a diagram indicating causal relationships among the specified variables. (Feel free to use the back page if needed). For example:

