Newton's Playground: How to use evidence centered design (ECD) to develop game-based assessment

Yoon Jeon Kim, Matthew Ventura, Valerie J. Shute, Russell Almond
Florida State University—1114 W. Call Street, Tallahassee, FL
Email: yjkim.fsu@gmail.com, mventura@fsu.edu, vshute@fsu.edu, ralmond@fsu.edu

Abstract: Good video games tend to elicit numerous behaviors that can be used as the basis for enormous amounts of data that can be used to inform competency states. This begs the question of how video games can be used as assessment tools. In this workshop, we will present an implementation of a game-based assessment in the video game *Newton's Playground.* As part of the activity, participants of the workshop will create problems that assess the competency of persistence.

Introduction

The main claim of researchers in the field of games and learning is that video games can facilitate learning because games provide a rich, engaging context that is conducive for learning to occur (Gee, 2003; Shaffer, Squire, Halverson, & Gee, 2005). Additionally, video games have many affordances for assessment (Gee & Shaffer, 2010; Shute, Ventura, Kim, & Wang, in press). Nevertheless, the notion of game-based assessment is still fairly new to the field of learning and assessment. To move forward with the idea of game-based assessment, we first need to understand how to marry the science of assessment development and the art of game design. Some of the questions that one needs to answer in the process of developing game-based assessment include: What does the traditional sense of item difficulty mean in a game-based assessment? How should difficult "items" in a game-based assessment look like? How should game designers develop in-game problems with varying difficulty? A powerful framework that can be used to tie game and assessment design principles together is evidence-centered design (Mislevy, Steinberg, & Almond, 2003). Evidence-centered design (ECD) has been the most pervasive assessment framework for game and simulation-based assessment, and has been used to develop multiple assessment systems (e.g., SimScientists, Cisco Packet Tracer). ECD offers a systematic approach to coherently aligning tasks or missions of the game that elicit evidence for the competencies of interest.

The proposed workshop aims to (a) review how ECD can be used as a framework that bridges game and assessment design principles, (b) re-think psychometric features of assessment (e.g., item difficulty) in the context of games, and (c) provide an opportunity for participants to create game-based assessments to measure persistence.

Evidence-Centered Design (ECD) for Game-Based Assessment

The primary purpose of an assessment is to collect information that will enable the assessor to make inferences about learners' competency states—what they know, believe, can do, and to what degree (Shute, 2009). ECD defines a framework that consists of three main theoretical models—competency, evidence, and task models that work in concert (Mislevy & Haertel, 2006; Mislevy, Steinberg, & Almond, 2003). When the ECD framework is coupled with the game design process, it allows game designers and assessors to (a) decide and define the claims to be made about learners' skills, knowledge, and other attributes, (b) identify what behaviors in the game constitute evidence of the claim, and (c) determine the nature and form of problems in the game that will elicit that evidence. Therefore, a good gamed-based assessment elicits behavior that bears evidence about key competencies, and it must also provide principled interpretations of that evidence in terms that suit the purpose of the assessment. The following section describes three primary models of ECD (i.e., competency, evidence, and task models), and their roles in designing game-based assessments.

<u>Competency Model</u>

A competency model (CM) is a structure of knowledge, skills, and other attributes to be assessed. Although ECD can work with a single variable CM, its strength resides in more complex multidimensional assessment, which is often the case in game-based assessment. Variables in the CM describe the set of knowledge and skills on which inferences are based (Almond & Mislevy, 1999). In game-based assessment, game designers and subject-matter experts need to determine, at the very early stage of the design, what skills and knowledge the game requires for players, and how those can be structured in a meaningful way. Some of the guiding questions to build a CM include:

"What kinds of skills do we want players to use to solve problems in the game?" "What attributes (e.g., persistence, curiosity) are relevant for players to succeed in the game"? Such questions are also commonly asked in game design processes.

Evidence Model

An evidence model (EM) expresses how the student's interactions with, and responses to, a given problem constitute evidence about CM variables. Also, an EM specifies what behaviors or performances should reveal those competencies. The EM indicates mathematical relationships between those behaviors and the CM variable(s). The task of specifying EMs is compatible with setting up rules of play in game design. Game designers often ask, "What do players need to do to move to a next level, and what counts as evidence to what extent?" Game and assessment designers need to ask such questions to construct EMs in game-based assessment.

Task Model

A task model (TM) describes what tasks or problems should be used to elicit behaviors defined in the evidence model. TM variables describe features of situations and tasks that will be used to elicit performance. A TM provides a framework for characterizing or constructing situations with which a student will interact to provide evidence about targeted aspects of competencies. The main purpose of tasks or problems is to elicit evidence (observable) about competencies (unobservable). In game-based assessment, TMs specify what players are expected to do and the features of problems or missions in the game environments with which players interact.

Stealth assessments in Newton's Playground

Stealth Assessment

Stealth assessment refers to ECD-based assessments that are woven directly and invisibly into the fabric of the learning environment (Shute, 2011). During game play, students naturally produce rich sequences of actions while performing complex tasks, drawing on the very skills or competencies that we want to assess (e.g., creativity, physics understanding). Evidence needed to assess the skills is thus provided by the players' interactions with the game itself (i.e., the processes of play).

Newton's Playground

In this workshop, we will share some examples from our development project of stealth assessments in a physics game called *Newton's Playground* (NP). NP is a computer game that emphasizes two-dimensional physics simulations, including gravity, mass, kinetic energy, and conservation of momentum. The goal of each problem in NP is to guide a green ball from a predetermined starting point to a balloon. Everything in the game obeys the basic rules of physics relating to gravity and Newton's three laws of motion. The primary way to move the ball is by drawing physical objects on the screen that "come to life" once the object is drawn. For example, in the "golf problem" as shown in Figure 1, the player must draw an object similar to a golf club to make it swing and hit the ball. Also, the player needs to draw a ramp-like device to direct the ball to the balloon. The speed of the swinging golf club is dependent on the size and mass of the club drawn and the angle from which the player drops it to swing. The ball will then fly at a certain speed, length, and trajectory. We want to point out that NP is inspired by a popular physics game called *Crayon Physics Deluxe*. In fact, NP has the identical core game mechanics with *Crayon Physics Deluxe* (i.e., drawing physical objects to create forces in a 2D environment). Our motivation to develop NP was to enable us to incorporate assessment mechanics seamlessly into the game environment

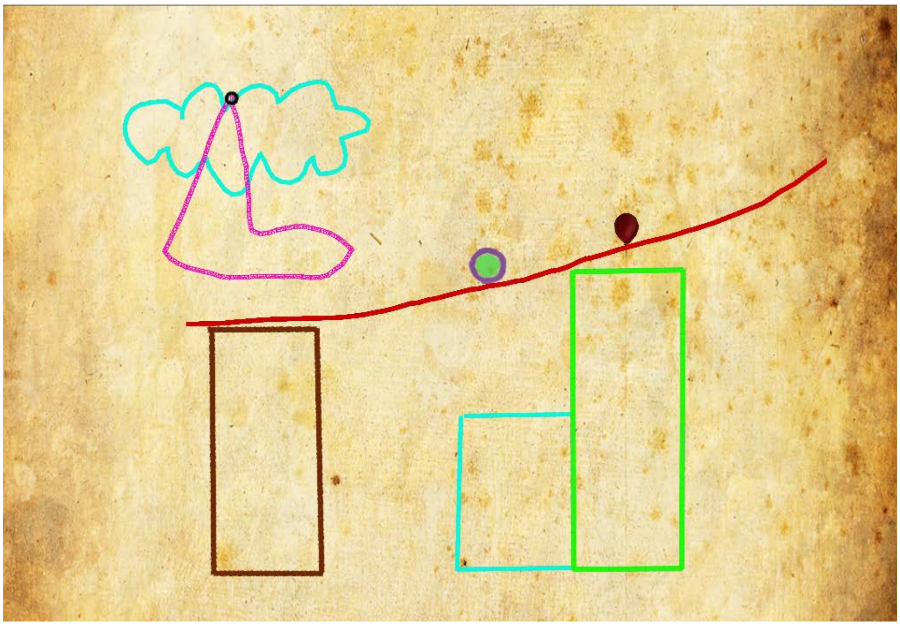

***Figure 1:* The Golf Problem**

Various problems in NP require the player to create devices such as levers, pendulums, and so forth to reach the ball to the balloon. Thus all solutions provided by the player can be categorized based on *agents of motion*—physical objects that generate a force or change the magnitude or direction of a force. The following is agents of motion that are used in the game and relevant physics principles:

1. *Ramp*: A ramp can be employed to change the direction of the motion of the ball (or another object). In some cases, other devices (like a pendulum or nudge), are needed to get the ball to start moving.
2. *Lever*: A seesaw or lever involves net torque. A lever rotates around a fixed point usually called a fulcrum or pivot point. An object residing on a lever gains potential energy as it is raised.
3. *Pendulum*: A swinging pendulum directs an impulse tangent to its direction of motion. The idealized pendulum is a specialized case of the physical pendulum for which the mass distribution helps determine the frequency. One can draw a physical pendulum in NP, and the motion will be determined by the mass distribution.
4. *Springboard*: A springboard (or diving board) stores elastic potential energy provided by a falling weight. Elastic potential energy becomes kinetic as the weight is released.
5. *Pin*: A pin allows the position of one body to be fixed in space. Like a nail, it supplies a force large enough to resist motion of the point it is attached to. Two pins hold a body immobile against a background.
6. *Rope*: Ropes generally transmit tension between objects. Ropes can also acts like trampolines, generating forces on objects by stretching the rope and then removing the force (by deleting objects) to produce upward momentum on the ball.
7. Nudge: An arrow in NP allows the player to gently nudge an object into motion.

Assessment Development

We identified three core competencies that can be assessed in NP: conscientiousness, creativity, and conceptual physics understanding. Conscientiousness (C) is a multi-faceted competency that commonly includes tendencies related to being attentive, hard-working, careful, detail-minded, reliable, organized, productive, and persistent (Noftle & Robins, 2007). Meta-analyses have linked conscientiousness with grades between r = .21 and .27, and the relationship is independent of intelligence (e.g., Noftle & Robins, 2007, Proporat, 2009). In this workshop we will focus on building tasks that will elicit behaviors (i.e., indicators) for one facet of conscientiousness, persistence. Persistence can be simply defined as willingness to work hard in spite of difficulty. Figure 2 displays the CM of conscientiousness as it is applied to NP. As can be seen, the facet persistence is highlighted and linked to behavioral indicators in NP gameplay.

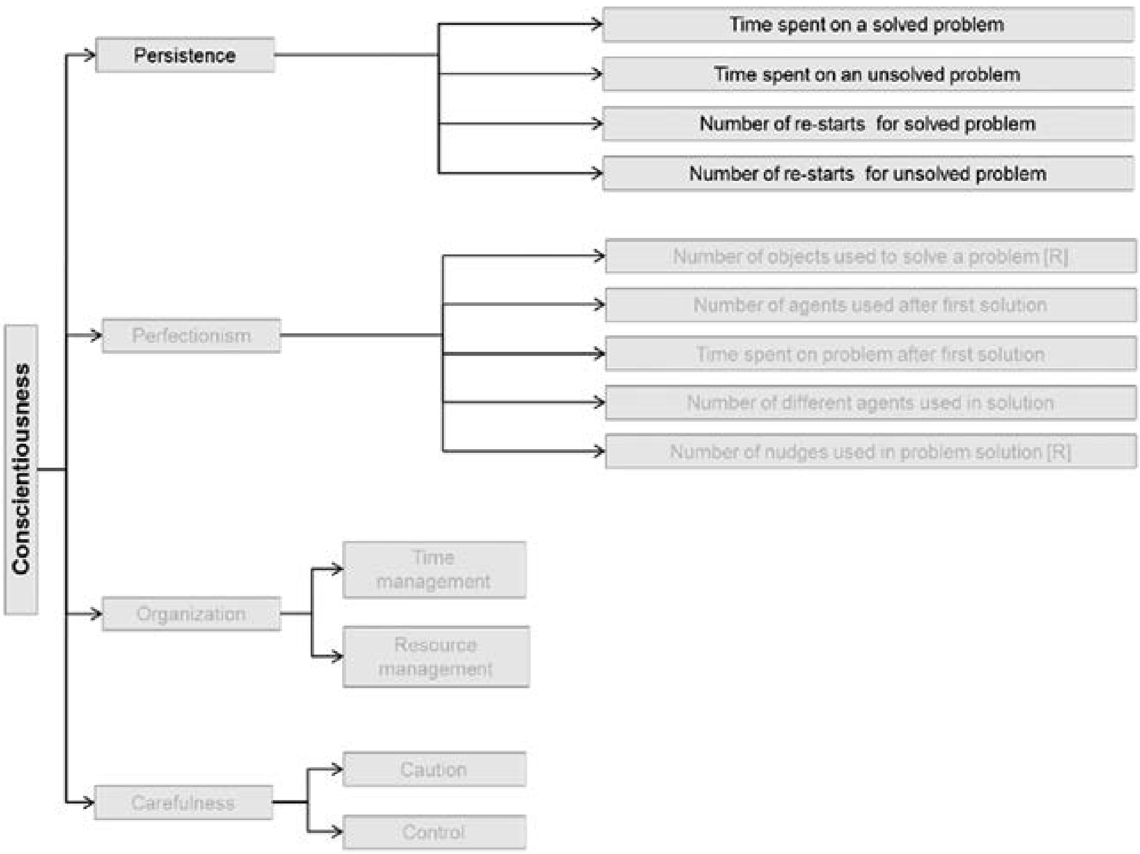

Figure 2: **A competency model of conscientiousness in Newton's Playground**

One of the challenges we are facing is figuring out how to ensure selected and created problems in the game are suitable for eliciting evidence for our three focal competencies. For example, assessing persistence is primarily based on seeing how long players spend or persist on difficult problems. Therefore, a good problem to assess persistence needs to have a "right level of difficulty" where the majority of players find it difficult yet interesting.

To address this issue we created difficulty rubrics for problems in order to systematically manipulate problem difficulty. This allows us to incrementally increase the difficulty of problems to ensure that students will eventually get to problems they will have trouble solving. Difficulty rubrics include:

1. *Relative location of ball to balloon*. If the balloon is positioned above the ball in a problem, this forces the player to use a lever, springboard, or pendulum to solve the problem (0-1 point).
2. *Obstacles*. This refers to the pathway between the ball and balloon. If the pathway is obstructed, this requires the player to project the ball in a very specific trajectory to obtain the balloon (0-2 points).
3. *Distinct agents of motion.* A NP problem may require one or two agents of motion to get the ball to the ball (0-1 point).
4. *Novelty.* This addresses whether a problem is novel relative to other problems played. Problem solution is not easily determined from experience with other problems (0-2 points).

Each problem is then evaluated under these rubrics to yield a difficulty score, and each of these rubrics contributes to the difficulty of the problem in a different way. Using these rubrics, a difficulty score can range from zero to six. For example, consider the problem cave story (Figure 3). This problem's difficulty score is 5 based on the rubrics. Thus the cave story can be a good problem to assess persistence since it will likely be unsolvable to some students.

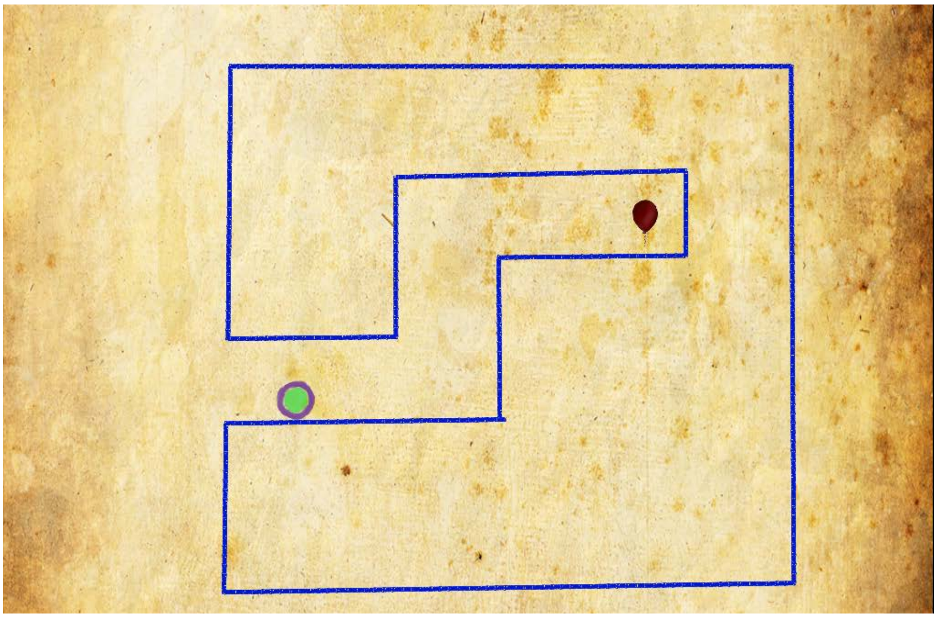

Figure 3: **Cave Story: A good problem for persistence assessment**

Agenda

Our workshop will be organized as follows:

Presentation (5 minutes)
We will share information about game-based assessment and discuss how ECD fits with game design principles. We will also discuss what item/task difficulty means in the game context, and how to create problems with varying difficulty in a game. Additionally, we will highlight some of the examples from our project in NP.

Introduction to Newton's Playground (5 minutes)
We will share information about Newton's Playground, and briefly demonstrate how to solve problems and create new problems using the game editor. Mainly we will focus on demonstrating how to create *agents of motion* that players can use in NP to move the ball. Some of the agents of motion include:

Develop, playtest, and revise Newton's Playground problems for persistence (40 minutes)
In small groups (of two or three), participants will create tasks of varying difficulty in Newton's Playground to assess persistence. Participants will playtest each other's tasks, and feedback for playtests will help participants revise their tasks.

Present-out and discussion (10 minutes)
Participants will present their tasks and how their tasks function as assessments for persistence. The whole group session will be followed by a reflection on the experience from the workshop.

Conclusion

We believe this workshop will benefit researchers, educators, and practitioners who want to use video games for learning and assessment. The focus of this workshop is designing problems in the game for assessment and support of learning. The idea of game-based assessment is still fairly new, and we believe this workshop will be a good venue to communicate ideas among participants with varying backgrounds. We further hope that this workshop will provide an opportunity for more people to understand that the fundamental philosophy of ECD is flexible enough to fit within diverse learning environments such as video games, and ECD can be a common language that game and assessment designers use to work collaboratively.

References

Almond, R. G., & Mislevy, R. J. (1999). Graphical models and computerized adaptive testing. *Applied Psychological Measurement, 23*(3), 223–237.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy.* New York, NY: Palgrave Macmillan.

Gee, J., & Shaffer, D. (2010). Looking where the light is bad: Video games and the future of assessment. *Edge: The Latest Information for the Education Practitioner, 6*(1), 3–19.

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 assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3–62.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology, 93,* 116-130.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135,* 322-338.

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. (2009). Simply assessment. *International Journal of Learning, and Media, 1*(2), 1-11. doi: 10.1162/ijlm.2009.0014

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte, NC:Information Age.

Shute, V. J., Ventura, M., Kim, Y. J., & Wang, L. (in press*).* Video games and learning. In W. G. Tierney, Z. Corwin, T. Fullerton, and G. Ragusa (Eds.), *Postsecondary Play: The Role of Games and Social Media in Higher Education*. Baltimore, MD: Johns Hopkins University Press.