Digital Games, Learning, and Stealth Assessment

Val Shute & Matthew Ventura (FSU)
July 21, 2011 (FSUS)
Project Overview

• We’re developing and evaluating assessments for some 21st Century competencies: creativity, persistence, and conceptual physics in a game called Crayon Physics Deluxe (CPD).

• CPD is a computer-based game that emphasizes 2D physics simulations, including gravity, mass, kinetic energy, and transfer of momentum.

• Data will be collected in CPD from players' interactions with the game. These data serve as the basis for the 3 “stealth assessments.”
Why use games as assessments?

1. Good games are engaging!
2. Games typically require a player to apply various competencies to succeed in the game (e.g., problem solving, persistence, creativity).
3. Games are ubiquitous. A survey of 1,102 youth (ages 12 -17) found that 97% play digital games: boys=99%; girls=94%. (Lenhart et al., 2008).
Assessment Framework
Evidence-Centered Design

(e.g., Mislevy, Steinberg, & Almond, 2003)
Assessment Design

**Competency Model**

What do you want to say about the person?

**Evidence Model**

What observations would provide best evidence for what you want to say?

**Task/Action Model**

What kinds of tasks let you make the necessary observations?
Design & Diagnosis

Assessment Models & Metrics

Competency

Evidence

Task/Action

Monitor and Diagnose Success

Competency Model: Organization of competencies & claims to be made about students, and current mastery estimates.

Evidence Model: Criteria or rubrics for evidence of claim (i.e., specific student performance data; observables).

Task/Action Model: A range of templates and parameters for task development to elicit evidence needed for the evidence model.
21st Century Competencies
Creativity is vital to society and culture re: advancing science, technology, humanities, and the arts (Dudek, 2003).

It’s also important to organizations and businesses re: its role in innovation and entrepreneurship. Nearly 50 years ago, Bruner (1962) called for society to embrace and foster children's creativity as a preparation for the future. That hasn’t yet happened.

We define creativity as the ability to derive novel solutions to various problems. Many problems in the world, and especially in the sciences, are often complex, ill-defined, and can’t be solved unilaterally. Tomorrow's innovative thinkers need to be able to think "outside the box" using creative strategies to solve problems.
Persistence

- We define **persistence as the motivation to work hard despite challenging conditions** (Peterson & Seligman, 2004).

- It’s been consistently found to predict academic achievement from preschool (Abe, 2005) to high school (Noftle & Robins, 2007, Proporat, 2009), to the postsecondary level (O’Conner & Paunonen, 2007) and adulthood (e.g., De Fruyt & Mervielde, 1996; Shiner, Masten, & Roberts, 2003).

- Meta-analyses have linked persistence with grades, and the relationship between persistence and grades is independent of intelligence (e.g., Noftle & Robins, 2007, Proporat, 2009).
• Understanding physics concepts is key to understanding the physical world, and understanding Newton's Laws of Motion is foundational, yet misconceptions abound (e.g., Clements, 1991; Gunstone & Watts, 1985; Halloun & Hestenes, 1985).

• In a study of flawed “common-sense” conceptions of force and motion shared among college physics students, Halloun and Hestenes (1985) determined that only 15% of their 478 experimental subjects demonstrated an accurate understanding of the relationship between unbalanced forces and acceleration (Newton’s 2nd law), while 61% confused the concepts of position, speed, and acceleration at least once.
FCI Assessment

- The Force Concept Inventory (FCI) will be used to test understanding of physics concepts (Hestenes & Halloun, 1995).
- The FCI includes 12 multiple-choice items: four questions on applying impulses to objects, four on interpreting kinematics (i.e., position, velocity, and acceleration) in dot trace representations, and four on the application of constant acceleration to objects.
- The FCI—recognized in the international physics education community as one of the best measures of academic conceptual understanding in physics—measures formalized conceptual understanding of Newtonian mechanics and kinematics.

A steel ball is attached to a string and is swung in a circular path in a horizontal plane as shown above. At the point P, the string suddenly breaks near the ball. If these events are observed from above, which path would the ball most closely follow after the string breaks?
Research Questions

1. Are the stealth assessments we develop in *Crayon Physics Deluxe* valid & reliable measures of our three focal competencies?

2. Do students learn conceptual physics from playing the game?

3. Are the ECD-based models we develop for Crayon Physics transferable to another game (e.g., *World of Goo*)?
Crayon Physics Deluxe

- Objective: guide a red ball from a starting point to a star. Everything obeys the basic rules of physics relating to gravity and Newton's three laws of motion.

- The player can nudge the ball (if the surface is flat) but the main way to move the ball is by drawing physical objects on the screen that "come to life" once the object is drawn.

- Various problems in CPD require the player to use catapults, pulleys, etc. to move the ball. Other problems require players to draw unique shapes that can be used to move the ball towards the star. In a number of cases the ball must go over a pit. If the ball falls into the pit, the player must start the problem over. Players can replay a problem as often as they like—even after successfully solving it.
Crayon Physics Deluxe

YouTube Video

(2 min., 47 sec.)
Year 1: Design Models & Pilot Problems

- We’ve begun creating 3 CMs and associated EMs (with relevant indicators).
- To define parameters of our task model, we’ll use existing problems in CPD and create new ones with the game's "level editor" tool. For our task models, we’ll identify problems that effectively elicit evidence for our 3 competencies.
- We’ll also pilot test ($n = 30-40$) the problems to determine if they're appropriate for our population and methodological requirements (e.g., adequate variability).
Year 2: Validity Research

• We’re running 2 studies to evaluate (a) validity & reliability of the stealth assessments, and (b) scalability of the CMs and EMs.

• 6th – 8th grade students (n = 120) will be comprise our sample. They’ll start by completing traditional tests (e.g., Torrance test for creativity; items from the International Personality Item Pool for persistence; and Force Concept Inventory for physics).

• We’ll also collect students’ GPA information.

• Students will then interact with 15 CPD problems over three one-hour sessions in the computer lab (i.e., roughly five problems per hour).
At the end of the last CPD session, students will complete a posttest of physics understanding (FCI) that will be compared to the physics pretest given at the beginning of the study.

Final competency estimates will be correlated with the scores from the traditional tests of the same competencies.

The results of the study will inform us as to the validity of the stealth assessments for the three focal competencies (creativity, persistence, and physics) and provide us with preliminary evidence for physics learning in CPD.
Year 2: Scalability Research

• We also plan to analyze the degree to which we can "plug 'n play" our stealth assessment models (persistence), from one game to another (e.g., World of Goo).

• We’ll modify the task models – specifying problem types in the second game that can be used to elicit the same evidence for persistence as in CPD (e.g., time spent replaying a problem already solved to improve the score, and so on).

• A subset of the students from Study 1 will be used in Study 2 (n = 80) to evaluate the effectiveness of the recycled stealth assessment. Students will interact with the second game over two 1-hr sessions. During each session, estimates of persistence will be updated based on student gameplay.
Year 2: Scalability Research

• After the sessions, Study 2 persistence estimates will be correlated to Study 1 persistence estimates and to scores from the traditional tests. This will provide information about the validity of the second-game assessment implemented in Study 2.

• Scalability (i.e., cost-benefit analysis) will be examined relative to the challenges (costs) associated with the time it takes to modify the models to suit the second game, and the effort needed to insert the models into the second game. We’ll monitor costs in terms of the time required for these efforts, and examine benefits relative to the validity of the competency estimates and learning.
Thank you!

Questions?

Email: vshute@fsu.edu
Website: http://www.myweb.fsu.edu/vshute