CRESST REPORT 773

Girlie C. Delacruz
Gregory K.W.K Chung
Eva L. Baker

VALIDITY EVIDENCE FOR GAMES AS ASSESSMENT ENVIRONMENTS

JULY, 2010

National Center for Research on Evaluation, Standards, and Student Testing

Graduate School of Education & Information Studies
UCLA | University of California, Los Angeles

Validity Evidence for Games as Assessment Environments

CRESST Report 773

Girlie C. Delacruz, Gregory K. W. K. Chung, & Eva L. Baker
CRESST/University of California, Los Angeles

July, 2010

National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Building, Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the National Research and Development Centers, PR/Award Number R305C080015, as administered by the U.S. Department of Education, Institute of Education Sciences.

The findings and opinions expressed in this report do not necessarily reflect the positions or policies of the National Research and Development Centers of the U.S. Department of Education, Institute of Education Sciences.

To cite from this report, please use the following as your APA reference:
Delacruz, G. C., Chung, G. K. W. K., & Baker, E. L. (2010). *Validity evidence for games as assessment environments* (CRESST Report 773). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

VALIDITY EVIDENCE FOR GAMES AS ASSESSMENT ENVIORNMENTS

Girlie C. Delacruz, Gregory K. W. K. Chung, & Eva L. Baker
CRESST/University of California, Los Angeles

Abstract

This study provides empirical evidence of a highly specific use of games in education—the assessment of the learner. Linear regressions were used to examine the predictive and convergent validity of a math game as assessment of mathematical understanding. Results indicate that prior knowledge significantly predicts game performance. Results also indicate that game performance significantly predicts posttest scores, even when controlling for prior knowledge. These results provide evidence that game performance taps into mathematical understanding.

Introduction

Games as assessment contexts

Games have long been attractive as learning environments given that games can entertain, motivate, and energize us. This report will address a highly specific use of games in education—the assessment of the learner. Games can be used as formative (in-the-process-of-learning) assessments, as well as for criterion trials, either to determine the level of performance of an individual or to gauge the speed and agility with which a learner acquires a new set of skills in an unfamiliar game environment (Baker & Delacruz, 2007; Gee, 2008). When designed properly, the underlying game engine can enable increases in challenge, complexity, and the cognitive demands required as the game progresses such that game play can be one form of assessment. Assessment is a process of drawing reasonable inferences about what a person knows by evaluating what they say or do in a given situation. However, it is insufficient to state that an assessment task is or is not valid. Rather, determining the validity of assessment tasks requires creating an argument that examines how well assessments answer the questions they purport to answer, as well as ensuring the data obtained provide the appropriate evidential basis for the claims made about students (American Educational Research Association, American Psychological Association, and National Council for Measurement in Education, *Standards for Educational and Psychological Testing,* 1999). In this study, we report findings that investigated the validity of a mathematics game as assessment of mathematical understanding by examining the relationship between mathematical knowledge and performance in the game.

Puppetman as an assessment context

The researchers at the National Center for the Research on Evaluation, Standards, and Student Testing (CRESST) along with game developers from the University of Southern California developed a mathematics game called *Puppetman*, which targets two pre-algebraic concepts: (a) defining a "unit" and (b) addition of rational numbers (integers and fractions). Specifically, game play in *Puppetman* focused on the idea that all rational numbers are defined relative to a single, unit quantity (e.g., a unit of count, measure, area, volume) and that rational numbers can be summed only if the unit quantities are identical. *Puppetman* is a puzzle game in which players need to determine the appropriate units to navigate from a starting point to a goal. Players build trampolines using coils, which determine how far *Puppetman* will bounce. The trampolines can bounce *Puppetman* left to right (Figure 1), or right, up, and down (Figure 2).

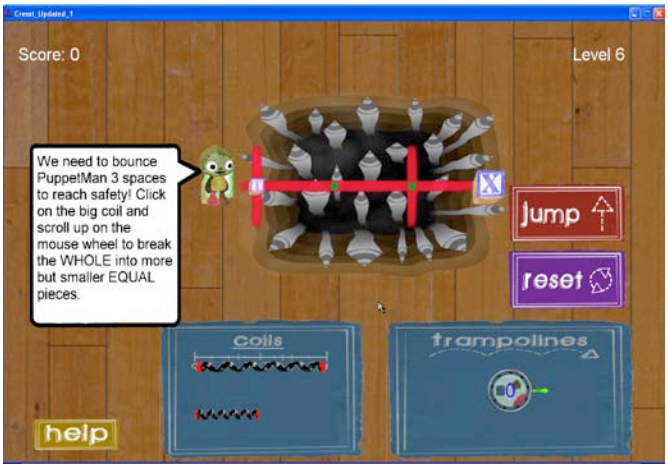

Figure 1. Screenshot of early level in *Puppetman*.

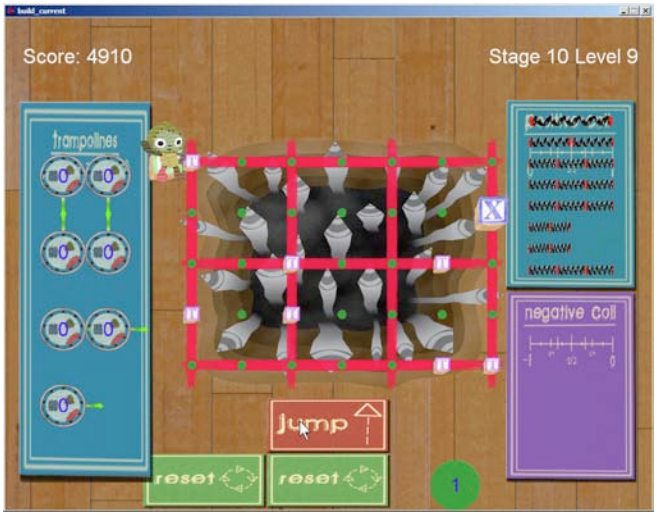

Figure 2. Screenshot of advanced level in *Puppetman.*

Method

Data were collected from 134 summer school high school students. A pretest was administered that included items comprised of adding fractions, determining the size of the unit in various graphical representations, and completing word problems. Students were given between 30 to 40 minutes to play the game. They then took a posttest, which included the same items on the pretest, some additional math items that incorporate game features from *Puppetman*, and a background questionnaire, which targeted attitudinal and interest information in both mathematics and games.

Analysis

In order to examine the validity of *Puppetman* as an assessment task, using a linear regression framework, we examined the predictive validity of pretest scores (i.e., prior knowledge) on game performance, and to obtain convergent evidence, we examined the predictive validity of game performance on posttest scores.

Puppetman was designed to get increasingly complex, with the latter levels requiring the most knowledge of mathematics to be successful. Thus, we used the last level attained as our metric for game performance.

We collected validity information by examining the relationship between various math outcomes scales: (a) pretest scores on pre-algebra items targeting rational number concepts (e.g., identifying numbers on a number line), (b) a smaller subscale of items on the pretest that directly relate to *Puppetman* content (e.g., symbolically adding fractions or identifying the size of a unit), and (c) and items on the posttest that comprised of both of the pretest scales above, as well as additional items that asked students to use the mathematics learned in the game to solve problems posed within the game context.

Results

Reliability analyses were conducted on three scales of math outcomes to ensure that the data of each scale had a unidimensional structure. First, the *pretest scale* was comprised of eight items on the pretest that targeted a range of conceptual understanding of fractions. These items asked students to translate graphical representations of fractions into its symbolic counterparts, identify fractions and define a unit on a number line. The Cronbach's alpha for this pretest scale was .63. Another scale was formed for four computational adding fraction items on the pretest. The Cronbach's alpha was for this scale was .73. Finally, the third scale was comprised of 21 items on the posttest. These included the same computational adding fraction items as the pretest, isomorphs of the symbolic items within the context of

the game, and other items that asked students to define a unit and represent fractions. The Cronbach's alpha for this scale was .88.

Predictive Validity Results

Descriptive statistics were obtained on the *pretest* items, pretest *adding fraction* items, and game performance. For the 134 students, the average pretest score was 6.26 (*SD =3.40*). The average pretest score on the adding fraction items was 2.21 (*SD =1.44*). The average last level attained in the game was 14.01 (*SD =3.19*).

A linear regression analysis indicated that math pretest scores significantly predicted game performance, $\beta = .546$, $t(133) = 8.234$, $p < .001$. Performance on the math pretest also explained a significant proportion of the variance in game performance, $R^2 = .34$, $F(1, 132) = 67.698$, $p < .001$. Performance on the adding fraction items also significantly predicted game performance, $\beta = 1.036$, $t(115) = 6.084$, $p < .001$. Performance on the math pretest also explained a significant proportion of game performance, $R^2 = .22$, $F(1, 132) = 37.02$, $p < .001$.

Convergent Validity Results

Descriptive statistics were obtained on performance on the posttest items and game performance (Note: three students were dropped from the original sample because they did not take the posttest). For the 131 students, the average posttest score was 13.05 (*SD =5.51*). The average last level attained in the game was 14.09 (*SD = 3.12*).

The linear regression analysis indicated that game performance significantly predicted performance on the posttest items, $\beta = .67$, $t(128) = 2.86$, $p < .05$. Game performance also explained a significant proportion of the variance in performance on the posttest items, even after controlling for pretest scores, $R^2 = .57$, $F(2, 128) = 86.38$, $p < .001$, with game performance explaining ***3% of the variance above and beyond*** pretest scores.

Conclusion

This study presented empirical evidence to support our claim that games can be valid assessment contexts. It is important to note that although space did not permit the presentation of survey results, responses to the survey indicate that many students did not perceive *Puppetman* to be a test and responded that they would like to play it at home and during school. While we present empirical evidence of validity of one game as an assessment context, this study demonstrates the potential for games to be valid assessments of understanding. We are currently analyzing the process data in the game play (e.g., time spent on each level, specific actions taken) as a formative assessment tool, to gain insight into the strategies that players employ while playing *Puppetman*.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Baker, E. L., & Delacruz, G. C. (2008). A conceptual framework for assessment of learning in games. In H. F. O'Neil & R. S. Perez (Eds.), *Computer games and team and individual learning* (pp. 21–38). Amsterdam, Netherlands: Elsevier.

Gee, J. (2008). Game-like learning: An example of situated learning and implications for opportunity to learn. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn* (pp. 200–221). New York, NY: Cambridge University Press.