CRESST REPORT 835

THE EFFECT OF IN-GAME ERRORS ON LEARNING OUTCOMES

OCTOBER, 2013

Deirdre Kerr

Gregory K.W.K. Chung

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

UCLA | Graduate School of Education & Information Studies

The Effect of In-Game Errors on Learning Outcomes

CRESST Report 835

Deirdre Kerr and Gregory K.W.K. Chung
CRESST/University of California, Los Angeles

October2013

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 Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Centers Program, PR/Award Number R305C080015.

The findings and opinions expressed here do not necessarily reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences, or the U.S. Department of Education.

To cite from this report, please use the following as your APA reference: Kerr, D. & Chung, G. K. W. K. (2013). *The Effect of In-Game Errors on Learning Outcomes* (CRESST Report 835). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

THE EFFECT OF IN-GAME ERRORS ON LEARNING OUTCOMES

Deirdre Kerr and Gregory K.W.K. Chung
CRESST/ University of California, Los Angeles

Abstract

Student mathematical errors are rarely random and often occur because students are applying procedures that they believe to be accurate. Traditional approaches often view such errors as indicators of students' failure to understand the construct in question, but some theorists view errors as opportunities for students to expand their mental model and create a deeper understanding of the construct. This study examines errors in an educational video game that are indicative of two specific misunderstandings students have about fractions (unitizing and partitioning) in order to determine whether the occurrence of those errors makes students more likely to learn from the game or more likely to be confused by the game. Analysis indicates that students who made unitizing errors were more likely to be confused by the game while students who made partitioning errors were more likely to learn from the game.

Introduction

Student mathematical errors are rarely random or capricious (Radatz, 1979). More often than not, student errors stem from organized strategies and rules that have sensible rationales (Bejar, 1984). Most errors are not accidental, but arise from ideas, rules, or procedures that make sense to the student and are perceived by the student as being accurate (Movshovitz-Hadar, Zaslavsky, & Inbar, 1987). These erroneous ideas, rules, or procedures frequently produce correct answers under specific circumstances, though they are at odds with mathematical concepts or reasoning and therefore are inaccurate in other circumstances (Babai, 2010).

Traditional approaches often view errors as failures on the part of the student (Bejar, 1984). For many theorists, errors are perceived as having a negative effect on learning because they increase frustration, reduce motivation, and produce stress, which leads to increased cognitive overload and the tendency to rely on old habits even if the student knows they are not productive (Frese & Altman, 1989). However, cognitive theorists and action theorists note that errors can provide information that the student would not otherwise encounter, thereby leading to a more comprehensive mental model of the construct (Frese & Zapf, 1994). Additionally, correcting an error can serve to improve future performance and broaden understanding by facilitating revisions to a previously faulty knowledge structure (Ohlsson, 1996).

Educational video games are a promising medium for addressing errors in this more formative manner for several reasons. First, because games are designed to challenge students and develop their skills, errors are expected and carry little or no negative stigma, (Pinelle, Wong, & Stach, 2008). Secondly, when students make errors during a game the software can provide immediate targeted feedback, which is seldom the case in standard educational environments.

Student errors in educational games should be identified and analyzed with the purpose of selecting, implementing, and evaluating the instructional procedures to be used in-game (Jitendra & Kameenui, 1996). Information about the effects, positive or negative, of specific errors on learning can subsequently be used to modify the presentation of content in the game or provide feedback to students when an error is committed. Further, analyzing student mistakes may provide useful suggestions for improving the teaching and learning process of educational video games (Fiori & Zuccheri, 2005), thereby boosting the degree to which the games encourage students to think about the targeted educational content (Fisch, 2005).

Purpose

This study analyzed two different types of errors (unitizing and partitioning) made in an educational video game to determine what effects those errors might have on student performance. In-game errors were analyzed using a multinomial logistic regression and pretest-posttest design to answer the following research questions:

1. Do errors (unitizing or partitioning) affect the likelihood that students will be confused by a mathematics educational game?
2. Do errors (unitizing or partitioning) affect the likelihood that students will learn from a mathematics educational game?

The unitizing and partitioning errors used in the analysis were identified through a data mining technique known as cluster analysis. Cluster analysis is a density estimation technique for identifying patterns within user actions reflecting differences in underlying attitudes, thought processes, or behaviors (Berkhin, 2006; Romero, Gonzalez, Ventura, del Jesus, & Herrera, 2009) through the analysis of either general correlations or sequential correlations (Bonchi et al., 2001). Cluster analysis is particularly appropriate for the analysis of log data because clustering is driven solely by the data at hand and is therefore ideal in instances in which little prior information is known (Jain, Murty, & Flynn, 1999). Cluster analysis is used to identify the latent dimensionality of a data set (Roussos, Stout, & Marden, 1998) and compress the data set into a manageable number of variables that are nontrivial, implicit, previously unknown, and potentially useful (Frawley, Piateski-Shapiro, & Matheus, 1992; Hand, Mannila, & Smyth, 2001; Vogt & Nagel, 1992).

Cluster analysis partitions actions into groups based on a matrix of inter-object similarities (James & McCulloch, 1990) by minimizing within-group distances compared to between-group distances so that actions classified as being in the same group are more similar to each other than they are to actions in other groups (Huang, 1998). Cluster analysis considers two actions to be similar if they are both performed by the same students. Actions are considered to be different if some students perform one of the actions and different students perform the other. Properly used, cluster analysis algorithms can perform the necessary pattern reduction and simplification process so that patterns in large data sets can be detected (Vogt & Nagel, 1992). In the case of log files from educational video games, the identified patterns, or clusters, take the form of the different solution strategies and error patterns manifested by the students as they attempt to solve each game level.

Study Design

The data used in this study comes from the log files generated by an educational video game (*Save Patch*) designed to teach addition of fractions (Chung et al., 2010). In *Save Patch*, students are asked to place coils on trampolines to help Patch bounce over obstacles and reach his home on the far side of the screen.

To correctly solve each level, students must place trampolines at various locations along a one- or two-dimensional grid as indicated by the three small 'T' boxes located on the grid (Figure 1). Students then select coils from the Positive Coil or Negative Coil resource bins on the right side of the screen and place the coils on the trampolines to make them bounce. Students can choose the values of the coils provided or break the coils into smaller pieces. While any size coil can be placed on the trampoline initially, subsequent coils can only be added to the trampoline if they are the same size (i.e., have the same denominator). The distance Patch will bounce is the sum of all coil values added to the trampoline. For instance, if a student places two 1/3 coils on a trampoline, Patch will bounce 2/3 of a unit. The direction Patch will bounce is indicated by the green arrow attached to each trampoline.

2012).

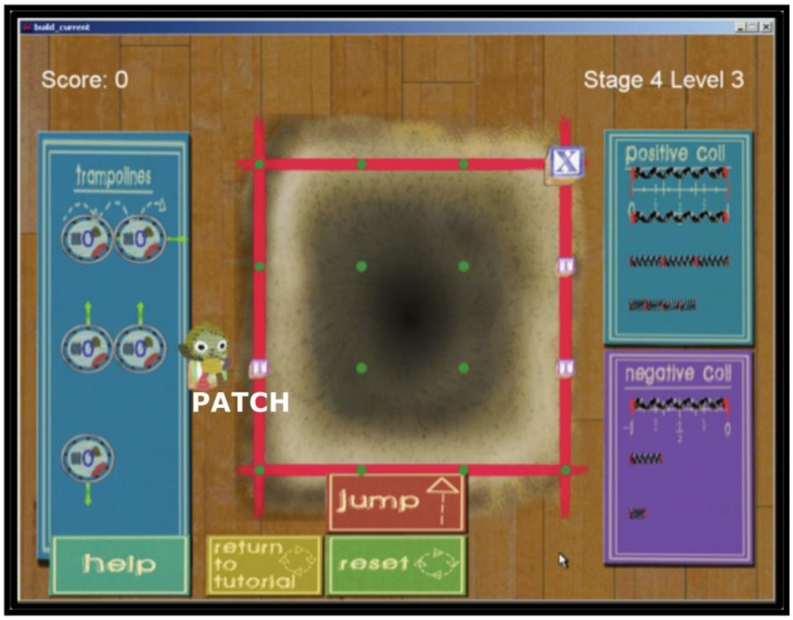

Figure 1. Screen shot of a level in Save Patch.

In *Save Patch*, one whole unit is always the distance between two red lines. Green dots indicate the size of the fractional pieces that should be used. In the level shown in Figure 1, the red lines indicate that the grid is one unit wide and one unit tall. The green dots inside the grid indicate that the grid is broken into thirds. To solve this level, a student would place a trampoline with an arrow pointed to the right on the first ‘T’ block, a trampoline with an arrow pointed up on the second ‘T’ block, and another trampoline with an arrow pointed up on the third ‘T’ block. Then the student would choose a whole unit coil from the Positive Coil resource bin and place it on the first trampoline, choose a 1/3 coil and place it on the second trampoline, and choose a 1/3 coil and place it on the third trampoline.

Our sample included 244 students (115 males, 115 females, and 14 students who did not report their gender) from a rural school district in southern California in either pre-algebra (63 students) or algebra (181 students) courses. Students in both courses took the pretest, played the game for approximately 40 minutes, and then took the posttest. Each action the students took in the game was logged automatically.

Methods

Cluster analysis using the fuzzy clustering algorithm “fanny” in R (R Development Core Team, 2010) was run to extract student error patterns. Two main mathematical errors were identified through the cluster analysis: unitizing errors and partitioning errors (Kerr & Chung, 2012).

Unitizing errors were made by students who were unable to correctly identify the unit in a given problem. Rather than using the hash marks that indicated units in the game, these students assumed that the length of the grid was always one unit, regardless of the number of hash marks present on the screen (Kerr, 2013). For example, in Figure 2, students who made a unitizing error would have identified the number line as representing one unit broken into eighths rather than representing two units broken into fourths.

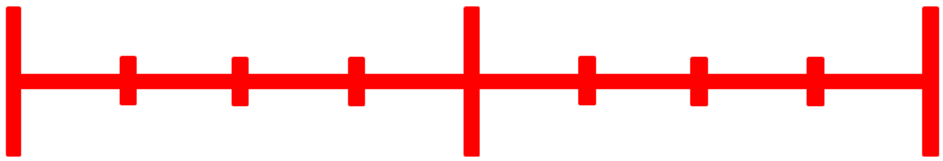

Figure 2. Number line representing two units broken into fourths.

Partitioning errors were made by students who were unable to correctly identify the fractional size the unit was broken into. Rather than counting the number of segments that the unit was broken into to determine the denominator, these students counted the markings separating the segments (Kerr, 2013). For example, again in Figure 2, students who made partitioning errors would have identified the number line as representing two units broken into thirds rather than two units broken into fourths.

To determine how these errors affected student performance, a multinomial logistic regression was run to determine whether the identified errors made students more likely to learn from the game (higher posttest score than pretest score) or more likely to be confused by the game (lower posttest score than pretest score). The variables representing in-game errors were binary variables indicating whether the student made unitizing errors in the game and whether the student made partitioning errors in the game. Additional variables controlled for in the regression included pretest score, positive self-reported feelings about the game, and low self-belief in math (as measured by items from Marsh, Hau, Artelt, Baumert, & Peschar, 2006).

All variables were calculated separately for students in pre-algebra and algebra, because these two groups of students differed in significant ways. Controlling for age, algebra students had significantly higher pretest scores ($p < .001$) and posttest scores ($p < .001$) and were marginally more likely to learn from the game ($p = .077$) than pre-algebra students.

Results

Not all variables in the initial multinomial logistic regression were significant predictors of the game's impact on students. Making partitioning errors, having positive feelings about the game, and having low self-belief in math were not significant predictors for pre-algebra students, perhaps due to the smaller sample size in that group. Therefore, these variables were dropped from the final analysis. With the reduced set of variables, the regression was significant at p = .007 and accurately categorized more of the students than would have occurred by chance.

The results showed that 91 students learned from the game, 61 students were confused by the game, and 92 students were not affected by the game. Pretest score was not a significant predictor of either confusion (p = .833) or learning (p = .609).

Table 1

Multinomial Logistic Regression Results

Variable	B	*SE*	Wald	*df*	Sig.	Exp(B)
Worse on posttest than pretest						
Intercept	-0.987	0.666	2.198	1	.138	
Low SBM (algebra)	0.999	0.443	5.078	1	.024	2.715
Positive view (algebra)	0.215	0.461	0.218	1	.641	1.240
Partitioning (algebra)	0.007	0.383	0.000	1	.986	1.007
Unitizing (algebra)	-0.230	1.094	0.044	1	.833	0.794
Unitizing (pre-algebra)	2.009	1.023	3.854	1	.050	7.453
Pretest score	0.016	0.038	0.184	1	.668	1.017
Better on posttest than pretest						
Intercept	0.260	0.573	0.205	1	.650	
Low SBM (algebra)	-0.358	0.485	0.544	1	.461	0.699
Positive view (algebra)	0.818	0.381	4.602	1	.032	2.265
Partitioning (algebra)	0.657	0.369	3.169	1	.075	1.928
Unitizing (algebra)	-3.258	1.209	7.258	1	.007	0.038
Unitizing (pre-algebra)	0.366	1.012	0.131	1	.718	1.442
Pretest score	-0.004	0.034	0.014	1	.907	0.996

Research Question 1

Do errors (unitizing or partitioning) affect the likelihood that students will be confused by a mathematics educational game?

As illustrated in Table 1, holding pretest constant, pre-algebra students were over seven times more likely to be confused by the game if they made unitizing errors in the game (p = .050, Exp(B) = 7.453). Holding pretest constant, algebra students were almost three times more likely to be confused by the game if they had low self-belief in math (p = .027, Exp(B) = 2.715).

Research Question 2

Do errors (unitizing or partitioning) affect the likelihood that students will learn from a mathematics educational game?

As illustrated in Table 1, holding pretest constant, algebra students were twice as likely to learn from the game if they felt positively about the game (p = .032, Exp(B) = 2.265) or made partitioning errors (p = .073, Exp(B) = 1.928), but were far less likely to learn from the game if they made unitizing errors (p = .007, Exp(B) = 0.038).

Significance

These results indicate that different types of errors (unitizing and partitioning) may have different effects on learning. While students who made unitizing errors were more likely to be confused by *Save Patch* and less likely to learn from the game, students who made partitioning errors were more likely to learn from the game. There are two possible reasons the errors might differ in their effects.

On one hand, the feedback given for partitioning errors in the game may have been superior to the feedback given for unitizing errors. If a student is unable to interpret the feedback they are given upon making that error, errors will be seen as a point of frustration rather than a learning opportunity (Frese & Altman, 1989). Effective feedback should help learners identify the aspects of the procedure that are relevant to their decision-making process, rather than simply notifying learners that they have made an error or providing learners with the correct response (Ohlsson, 1996). Our feedback for the partitioning error may have followed this format more closely than our feedback for the unitizing error.

On the other hand, unitizing errors and partitioning errors may be vastly different types of errors. Researchers have identified two different kinds of errors: slips and mistakes. Slips are failures in execution, whereas mistakes are failures to adequately grasp the concept (Reason, 1995). Slips are often caused by internal or external distractions (Zhao & Olivera, 2006), and the action is not performed as intended (Norman, 1981). Mistakes, on the other hand, arise from misconceptions or misapplications of rules or procedures (Reason, 1995). Novices commit more mistakes, while experts commit more slips (Frese & Altman, 1989). Therefore it’s possible that unitizing errors are mistakes while partitioning errors are slips, and mistakes have negative effects on learning whereas slips do not.

Despite the ambiguity of interpretation, this analysis provided us with results that we can use to improve the design of our game so that it is a more effective instructional platform. It may improve student learning if we change the feedback given to students when they make unitizing errors to point out how information about the unit size is used in determining the denominator of a given representation. It may also improve student learning if we provide distracters that make partitioning errors attractive and remove distracters that make unitizing errors attractive. We are currently in the process of revising the game, and hope that the results from our upcoming analyses provide more insight into the effect of unitizing and partitioning errors on student learning.

References

Babai, R. (2010). Piagetian cognitive level and the tendency to use intuitive rules when solving comparison tasks. *International Journal of Science and Mathematics Education*, *8*(2), 203-221.

Bejar, I. I. (1984). Educational diagnostic assessment. *Journal of Educational Measurement*, *21*(2), 175-189.

Berkhin, R. (2006). A survey of clustering data mining techniques. In J. Kogan, C. Nicholas, & M. Teboulle (Eds.), *Grouping multidimensional data* (pp. 25-72). The Netherlands: Springer.

Bonchi, F., Giannoti, F., Gozzi, C., Manco, G., Nanni, M., Pedreschi, D., ..., Ruggieri, S. (2001). Web log data warehouses and mining for intelligent web caching. *Data & Knowledge Engineering*, *39*, 165-189.

Chung, G. K. W. K., Baker, E. L., Vendlinski, T. P., Buschang, R. E., Delacruz, G. C., ..., Bittick, S. J. (2010, April). Testing instructional design variations in a prototype math game. In R. Atkinson (Chair), *Current perspectives from three national R&D centers focused on game-based learning: Issues in learning, instruction, assessment, and game design*. Structured poster session at the annual meeting of the American Educational Research Association, Denver, CO.

Fiori, C., & Zuccheri, L. (2005). An experimental research on error patterns in written subtraction. *Educational Studies in Mathematics*, *60*(3), 323-331.

Fisch, S. M. (2005). Making educational computer games "educational." *Proceedings of the 4th International Conference for Interaction Design and Children*, 56-61.

Frawley, W. J., Piateski-Shapiro, G., & Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI Magazine*, *13*(3), 57-70.

Frese, M., & Altman, A. (1989). The treatment of errors in learning and training. In L. Bainbridge & S.A. Quintanilla (Eds.), *Developing skills with information technology* (pp. 65-86). Chichester, UK: Wiley.

Frese, M., & Zapf, D. (1994). Action as the core of work psychology: A German approach. In H. C. Triandis, M. D. Dunnette, & L. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 4, pp. 271-340). Palo Alto, CA: Consulting Psychologists Press.

Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Cambridge, MA: The MIT Press.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, *2*, 283-304.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys*, *31*, 264-323.

James, F., & McCulloch, C. (1990). Multivariate analysis in ecology and systematic: Panacea or Pandora's box? *Annual Review of Ecology and Systematics*, *21*, 129-166.

Jitendra, A., & Kameenui, E. J. (1996). Experts' and novices' error patterns in solving part-whole Mathematical word problems. *Journal of Educational Research, 90*(1), 42-51.

Kerr, D. (2013). Student perception of math content in an educational video game. In *Cognition and Assessment SIG Poster Session/Business Meeting*. Poster session at the 2013 annual conference of the American Educational Research Association, San Francisco, CA.

Kerr, D. & Chung, G. K. W. K. (2012). Identifying key features of student performance in educational video games and simulations through cluster analysis. *Journal of Educational Data Mining*, *4*, 144-182.

Marsh, H. W., Hau, K., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, *6*, 311–360.

Movshovitz-Hadar, N., Zaslavsky, O., & Inbar, S. (1987). An empirical classification model for errors in High School Mathematics. *Journal for Research in Mathematics Education, 18*(1), 3-14.

Norman, D. A. (1981). Categorization of action slips. *Psychological Review*, *88*, 1-15.

Ohlsson, S. (1996). Learning from error and the design of task environments. *International Journal of Educational Research*, *25*(5), 419-448.

Pinelle, D., Wong, N., & Stach, T. (2008). Heuristic evaluation for games: usability principles for video game design. *Proceedings of the 26th annual SIGCHI Conference on Human Factors in Computing Systems*, 1453-1462.

R Development Core Team. (2010). R: A Language and Environment for Statistical Computing [Computer software]. Retrieved from http://www.R-project.org

Radatz, H. (1979). Error analysis in mathematics education. *Journal for Research in Mathematics Education*, *10*(3), 163-172.

Reason, J. T. (1995). Understanding adverse events: human factors. *Quality in Health Care*, *4*, 80-89.

Romero, C., Gonzalez, P., Ventura, S., del Jesus, M. J., & Herrera, F. (2009). Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data. *Expert Systems with Applications*, *39*, 1632-1644.

Roussos, L., Stout, W., & Marden, J. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement*, *35*(1), 1-30.

Vogt, W., & Nagel, D. (1992). Cluster analysis in diagnosis. *Clinical Chemistry*, *38*(2), 182-198.

Zhao, B., & Olivera, F. (2006). Error reporting in organizations. *Academy of Management Review*, *31*, 1012-1030.