CRESST REPORT 819

THE MEDIATION EFFECT OF IN-GAME PERFORMANCE BETWEEN PRIOR KNOWLEDGE AND POSTTEST SCORE

SEPTEMBER, 2012

Deirdre Kerr

Gregory K.W.K. Chung

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

UCLA | Graduate School of Education & Information Studies

The Mediation Effect of In-Game Performance Between Prior Knowledge and Posttest Score

CRESST Report 819

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

September 2012

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. (2012). *The mediation effect of in-game performance between prior knowledge and posttest score* (CRESST Report 819). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

THE MEDIATION EFFECT OF IN-GAME PERFORMANCE BETWEEN PRIOR KNOWLEDGE AND POSTTEST SCORE[1]

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

Abstract

Though video games are commonly considered to hold great potential as learning environments, their effectiveness as a teaching tool has yet to be determined. One reason for this is that researchers often run into the problem of multicollinearity between prior knowledge, in-game performance, and posttest scores, thereby making the determination of the amount of learning attributable to the game difficult. This study uses tests for mediation effects to determine the true relationship between in-game performance and posttest performance, determining that in this case in-game performance is a perfect mediator of prior knowledge on posttest score.

Introduction

Many researchers believe video games hold great potential as learning environments because good principles of learning are built into good video game design (Gee, 2004), games are a lucrative, popular, and motivating medium (Squire, 2003), they allow instructional activities that are not possible in a traditional environment (Philpot et al., 2005), and they support multiple learning styles (Becker, 2005). However, despite broad agreement that educational video games could be powerful teaching tools, there is currently no consensus on whether students actually learn the intended content from the educational games they play (Tobias et al., 2011): the question of whether or not educational video games work as instructional environments is still unanswered.

One reason for this lack of consensus is that measuring the amount of learning attributable to a given educational video game is difficult. One of the most basic problems educational researchers face in this regard is that of multicollinearity, particularly between prior knowledge,

[1]This research was originally published as a paper in the Proceedings of the International Association of Science and Technology for Development (IASTED).

The APA citation for the original paper is: Kerr, D. & Chung, G. K. W. K. (2011). The mediation effect of in-game performance between prior knowledge and posttest score. In J. Matuga (Eds.), *Proceedings of the IASTED International Conference on Technology for Education (TE 2011)* (pp. 122-128). Anaheim, CA: ACTA Press. doi: 10.2316/P.2011.754-046.

We express our thanks to the IASTED for allowing us to republish this paper as a CRESST report.

in-game performance, and posttest score. Multi-collinearity arises when the independent variables in a regression model do not meet the requirement that they be independent of each other, posing a serious threat to the validity of the regression (Farrar & Glauber, 1967). This occurs when two independent variables are highly correlated with each other and both are highly correlated with the dependent variable. In such cases, the question for the researcher is which of several models accurately reflects the true relationship between these three variables.

For example, if Variable 1 is an independent variable (i.e., prior knowledge), Variable 2 is another independent variable (i.e., in-game performance) that could theoretically be related to both Variable 1 and Variable 3, and Variable 3 is a dependent variable (i.e., posttest score), there are four possible models of the true relationship between these three inter-correlated variables (see Figure 1).

- If Model 1 is the true representation of the relationship between these three variables, then Variable 1 (prior knowledge) predicts both Variable 2 (in-game performance) and Variable 3 (posttest score), and the correlation between Variable 2 (in-game performance) and Variable 3 (posttest score) is due solely to their shared relationship with Variable 1 (prior knowledge).
- If Model 2 is the true representation, then Variable 1 (prior knowledge) predicts Variable 3 (posttest score) and Variable 2 (in-game performance) predicts Variable 3 (posttest score), and the correlation between Variable 1 (prior knowledge) and Variable 2 (in-game performance) is due solely to their shared relationship with Variable 3 (posttest score).
- If Model 3 is the true representation, then Variable 1 (prior knowledge) predicts Variable 2 (in-game performance) and Variable 2 (in-game performance) predicts Variable 3 (posttest score), and the correlation between Variable 1 (prior knowledge) and Variable 3 (posttest score) is due solely to their shared relationship with Variable 2 (in-game performance).
- If Model 4 is the true representation, then Variable 1 (prior knowledge) predicts Variable 2 (in-game performance) and Variable 3 (posttest score), but Variable 2 (in-game performance) has an additional effect on Variable 3 (posttest score) beyond that of Variable 1 (prior knowledge) alone.

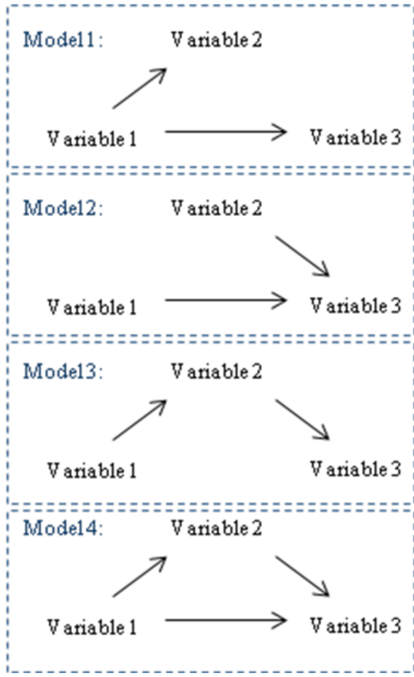

Figure 1. Four models of the possible relationships between three inter-correlated variables.

Each of these models carries different implications for researchers. For instance, if Model 1 is the true representation of the relationship between the three variables, then Variable 2 (in-game performance) is actually a second dependent variable rather than a second independent variable, and the strength of the relationship should be tested in two separate regressions: one regression of Variable 1 (prior knowledge) on Variable 2 (in-game performance) and one regression of Variable 1 (prior knowledge) on Variable 3 (posttest score). However, if Model 2 is the true representation then Variable 1 (prior knowledge) and Variable 2 (in-game performance) either measure similar traits or two different aspects of the same trait, and in order to investigate the strength of the relationship, an overall variable of the values of both Variable 1 (prior knowledge) and Variable 2 (in-game performance) should be computed and regressed on Variable 3 (posttest score). Finally, if Model 3 or Model 4 is the true representation then Variable 2 (in-game performance) is considered to be a mediator of Variable 1 (prior knowledge) on Variable 3 (posttest score).

A mediator is defined as a variable which acts as the mechanism through which an independent variable is able to influence a dependent variable (Baron & Kenny, 1986). In other words, all or part of the relationship between an independent variable and a dependent variable is explained by a second independent variable. If all of the relationship between the independent variable (prior knowledge) and the dependent variable (posttest score) is explained by the second independent variable (in-game performance), then Model 3 is the true representation and the second independent variable (in-game performance) is considered to be a perfect mediator. If only part of the relationship is explained by the second independent variable, then Model 4 is the true representation and the mediator is imperfect.

This study seeks to determine which of these four models represents the true relationship between prior knowledge, in-game performance, and posttest score in order to determine if in-game performance is a dependent variable (Model 1), a second measure of prior knowledge (Model 2), or a perfect (Model 3) or imperfect (Model 4) mediator of the effect of prior knowledge on posttest score.

Related Work

To our knowledge, no other researchers have performed a mediation study to determine the relationship between prior knowledge, in-game performance, and posttest scores. Rather, most studies on the effectiveness of educational video games or simulations simply report whether or not test scores after playing the game (posttest scores) are significantly higher than before playing the game (prior knowledge), using either an ANCOVA (Dinn & Calao, 2001; Kebritchi et al., 2010; Laffey et al., 2003; Serrano & Anderson, 2004; Sung et al., 2008), or a repeated measures ANOVA (Ke, 2008; Warren et al., 2008).

Only one study we came across mentioned that there might be a more complicated relationship between game play, prior knowledge, and posttest score (Virvou et al., 2005), and they did not investigate the phenomenon beyond stating that the amount of learning occurring in the game appeared to be related to level of prior knowledge.

Study Design

Sample

The data used in this study come from 17 students (11 females, 5 males, and 1 not stated) in a medium sized urban school district who played *Save Patch* for three days as part of a summer school program. 53% of students spoke a language other than English at home at least half of the time, 47% of the students were Latino/a, and 24% were African-American. All students had been sixth graders in the preceding school year. The average math grade on their last report card was a B-.

Game Design

Save Patch is an educational video game designed by the National Center for Research on Evaluation, Standards, and Student Testing at the University of California, Los Angeles and the University of Southern California's Game Innovation Lab to teach addition of fractions (Chung et al., 2010).

In this game, students are required to apply concepts underlying rational number addition to help the game character Patch safely navigate the path to treasure while collecting keys and coins and avoiding traps. To correctly solve each level, students must choose a path through a one-dimensional or two-dimensional grid and place the correct amount of unit fractions (i.e., fractions wherein the numerator is one) on each sign along that path. The distance Patch will travel is the sum of all unit fractions placed on the sign. For instance, if a student placed three 1/4 ropes on a sign, Patch would travel 3/4 of a unit.

In *Save Patch*, one whole unit is always represented by the space between the large gray posts on the dirt path. Smaller red posts, either on the path or in the grass, indicate the size of fractional pieces that should be used. Students can break ropes in the Path Options section into smaller fractions (i.e., a 1/2 piece can be broken into two 1/4 pieces) or combine them into larger pieces. Any sized rope can be placed on a sign initially. However, once a rope has been placed on a sign, only other ropes of the same size (i.e., same denominator) can subsequently be placed on the sign. See Figure 2 for an example level of *Save Patch*.

Figure 2. Example level from Save Patch.

The representation and game mechanics were specifically designed so that the actions a student took in the game would reflect the mathematical decisions being made, and the results of those actions would be consistent with the underlying mathematical concepts (Vendlinski et al., 2010).

Additionally, the different stages of the game were designed to represent the different steps students move through as they come to understand the addition of fractions. Therefore, *Save Patch* begins with the addition of whole units. Stage Two of the game extends addition to unit fractions (wherein the numerator of the fraction is 1, i.e., 1/1, 1/2, 1/3, etc.). Stage Three asks students to solve problems with both whole units and unit fractions on the same grid. Stage Four asks the students to solve problems wherein the whole unit distance crosses a unit marker (i.e., from 1/2 to 3/2). Stage Five asks students to solve problems wherein the answer lies somewhere between a unit fraction and a whole unit (i.e., 3/4). Finally, Stage Six extends the problems to those involving distances larger than a whole unit (i.e., 6/4).

Procedures and Methods

Study Procedures

The data in this study come from a summer school math class wherein educational video games were used to supplement curriculum. At the beginning of the course students were given a pretest to determine their overall level of mathematical knowledge of sixth grade concepts. Thirteen of the items on this test involved the addition of fractions. Student scores on those items constitute the prior knowledge measure for this study. Additional items on the test covered the multiplication and division of fractions and the calculation of rates. Those items were used to measure prior knowledge for games in other content areas, and were not used in the prior knowledge score for the addition of fractions.

Approximately one week later students began playing *Save Patch*. Each of the seventeen students played *Save Patch* until they had completed the criterion levels in the game. Immediately upon completion of those levels, students were given a posttest. That posttest included some of the items from the pretest as well as some additional items. All items related to the addition of fractions. *Save Patch* was the only game students played in this content area. Though the students also played games on multiplication and division of fractions and the calculation of rates, those games took place after the posttest and are immaterial to this study.

Students took two to four days to reach the criterion levels in *Save Patch*. Their mean prior knowledge score was 4.97, with a standard deviation of 2.51 and a range of 1 to 11 on a 13-point scale. Their mean in-game performance was 5.47, with a standard deviation of 2.90 and a range of 0 to 12 on a 12-point scale. Their mean posttest score was 4.76, with a standard deviation of 2.02 and a range of 2 to 8 on a 13-point scale.

Measuring In-Game Performance

In educational video games the process data used to identify key features of student performance are captured in the log files that are automatically generated by the game as students play. These files store complete student answers to the problems given in the game, including strategies and mistakes (Merceron & Yacef, 2004) thereby letting the researcher record the learning behavior of students as they play the game (Romero & Ventura, 2007). However, log files contain large quantities of noisy data and can be very difficult to analyze (Romero et al., 2009).

Identifying Student Strategies

One method of analyzing log data is 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; Bonchi et al., 2001) through the analysis of general or sequential correlations (Romero et al., 2009). It is particularly appropriate for the analysis of log data, as cluster analysis is driven solely by the data at hand and is therefore ideal in situations in which little prior information is known about the underlying structure of the data (Jain et al., 1999).

Cluster analysis partitions actions into groups on the basis of 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). Two actions will be considered to be similar by the cluster analysis if they are both performed by the same students. Two actions will be considered to be different from each other if some students perform one of the actions and different students perform the other action.

We used fuzzy clustering in R (Development Core Team, 2010) to identify different strategies students used in their attempts to solve each level of the game (Kerr et al., 2011). Some of those strategies were valid solution strategies, while others were identifiable error patterns. For instance, some students saw the entire grid as representing one whole unit, regardless of how many units were actually represented. In the level shown in Figure 2, this would result in a student attempting to solve the level using sixths rather than thirds. Other students recognized the unit correctly, but had difficulty determining the denominator. Frequently those students counted the red posts to determine the denominator rather than counting the spaces between the red posts. In the level shown in Figure 2, this would result in a student attempting to solve the level using halves rather than thirds.

Coding In-Game Performance

The results of the cluster analysis were used to code student in-game performance. Every student's initial attempt at each level was coded as being either a solution strategy or error pattern. We chose to examine initial attempts rather than all attempts because the first attempt a student makes is likely to be more indicative of the strategies they are using than later attempts, which will be at least partially based on the results of their previous attempts.

After all initial attempts were coded, each student was given an overall score for each stage in the game. If initial attempts for all levels were error patterns, then the student was given a 0 and was considered not to have learned the material. If initial attempts for the first few levels in a stage were error patterns and initial attempts for later levels in a stage were solution strategies, then the student was given a 1 and was considered to have learned the material. If initial attempts for all or nearly all levels in a stage were solution strategies, then the student was given a 2 and was considered to have known the material.

These scores were added up for each student across all six stages in the game, leading to a score between 0 and 12 for each student. This variable was then used as the measure of in-game performance in all subsequent analyses.

Testing for Mediation

In order to determine if in-game performance is a mediator of prior knowledge on posttest score, a series of three regressions must be run (Barron & Kenny, 1986). The first regression tests the path between Variable 1 (prior knowledge) and Variable 2 (in-game performance). If prior knowledge is not a significant predictor of in-game performance, then Model 2 (see Figure 1) is the true representation of the relationship and in-game performance is not a mediator.

The second regression tests the path between Variable 1 (prior knowledge) and Variable 3 (posttest score). The third regression tests the path between Variable 2 (in-game performance) and Variable 3 (posttest score) and retests the path between Variable 1 (prior knowledge) and Variable 3 (posttest score) when accounting for Variable 2 (in-game performance) by using both Variable 1 (prior knowledge) and Variable 2 (in-game performance) as independent variables. If in-game performance is not a significant predictor of posttest score controlling for prior knowledge in the third regression, then Model 1 is the true relationship and in-game performance is not a mediator of prior knowledge on posttest score.

However, if prior knowledge is a significant predictor of posttest score in the second regression and in-game performance is a significant predictor of posttest score in the third regression, then in-game performance is a mediator of prior knowledge on posttest score. If prior knowledge is no longer a significant predictor of posttest score in the third regression, then Model 3 is the true relationship and in-game performance is a perfect mediator of prior knowledge on posttest score. If, on the other hand, prior knowledge is still a predictor of posttest score in the third regression, but the relationship is less strong than in the second regression, then Model 4 is the true relationship and in-game performance is an imperfect mediator of prior knowledge on posttest score.

Results

The results of the mediation analysis are summarized in Table 1. For each of the three regressions in the mediation analysis, we have reported the percentage of the variance in the dependent variable that is explained by the independent variable or variables (Explained Variance), the amount of change in the dependent variable for every one unit change in the independent variable (Standardized Beta), and the *p*-value of the standardized beta coefficient (Significance).

Table 1

Regression Results

Regression	Independent variable	Dependent variable	Explained variance	Standardized Beta	Significance
#1	Prior knowledge	In-game performance	38%	.617	.008
#2	Prior knowledge	Posttest score	18%	.425	.089
#3	Prior knowledge In-game performance	Posttest score	60%	-.080 .819	.715 .002

In the first regression, prior knowledge was a significant predictor of in-game performance ($p = .008$), explaining 38% of the variance and having a standardized beta coefficient of .617. This indicates that a student's performance in the game depends on how much prior knowledge they have regarding fractions and rules out Model 2 as the true representation of the relationship (see Figure 1).

In the second regression, prior knowledge was a moderately significant predictor of posttest score ($p = .089$), explaining 18% of the variance and having a standardized beta coefficient of .425. This indicates that a student's performance on the posttest depends at least in part on how much prior knowledge they have regarding fractions.

In the third regression, in-game performance was a significant predictor of posttest score ($p = .002$). This indicates that a student's performance on the posttest depends on their in-game performance and rules out Model 1 as the true representation of the relationship.

Additionally, when in-game performance was accounted for in the third regression, prior knowledge was no longer a significant predictor of posttest score ($p = .715$) and the standardized beta coefficient for prior knowledge fell to effectively zero (-.080) while the standardized beta coefficient for in-game performance was .819. Furthermore, accounting for in-game performance in the regression resulted in an increase in the percentage of variance explained from 18% in the second regression to 60% in the third regression. This indicates that the effect of prior knowledge on posttest scores is indirect (through in-game performance) rather than direct, and therefore that in-game performance is a mediator of prior knowledge on posttest score. Since the relationship between prior knowledge and posttest score dropped to effectively zero, Model 3 appears to be the true representation of the relationship between the three variables, indicating that in-game performance is a perfect mediator of knowledge on posttest score.

The mediation effect of in-game performance between prior knowledge and posttest score is summarized in Figure 3. Figure 3 is the standard representation of a mediation study wherein the arrows indicate the relationships between variables, the values are standardized beta coefficients, and *p*-values are indicated with asterisks. The value in the parentheses (-.080) is the effect of prior knowledge on posttest score when controlling for in-game performance, while the value to the left of the parentheses (.425) is the effect of prior knowledge on posttest score before accounting for in-game performance.

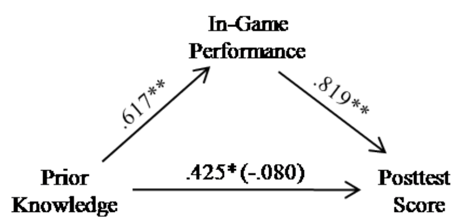

Figure 3. Mediating effect of in-game performance.
* = *p*-value < = .10 ** = *p*-value < = .05

If the variable in question (in-game performance) is a mediator, the value in the parentheses (-.080) will be lower than the value to the left of the parentheses (.425). If the variable is a perfect mediator, then the value in the parentheses will no longer be significant (indicated by the lack of an asterisk). Since -.080 is a much smaller value than .425, in-game performance is a mediator of the effect of prior knowledge on posttest score. Since -.080 is not significant—in-game performance is a perfect mediator.

Conclusion

The results of the mediation analysis suggest that prior knowledge of the subject area determines how well a student will perform in an educational video game, and that how well a student performs in an educational video game determines how well they will score on a posttest.

Additionally the results suggest that, when accounting for in-game performance, prior knowledge in the subject area does not determine how well a student will score on a posttest.

This indicates, at least in this sample of students playing this particular game, that learning does occur in games and that games can be effective teaching tools.

However, the sample size for this study was very small (17) and the findings would have to be replicated with a larger sample before any strong statements about the instructional potential of educational video games could be made. It would also be interesting to see if other games designed in a similar fashion have similar results.

Additionally, the strong relationship between prior knowledge and in-game performance suggests that our game is not successful in teaching students with low levels of prior knowledge of fractions. It might be useful in the future to try to reduce the strength of this relationship through targeted feedback or some other modification so that the chances of multicollinearity occurring in the first place could be reduced.

References

Baron, R.M., & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations, *Journal of Personality and Social Psychology*, *51*(6), 1173-1182.

Becker, K. (2005). *Games and learning styles*. Paper presented at the IASTED International Conference on Education and Technology, Calgary, Alberta, Canada.

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

Bonchi, F., Giannoti, F., Gozzi, C., Manco, G., Nanni, M. Pedreski, D., Renso, C., & 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., Michiuye, J.K., & Bittick, S.J. (2010, May). *Current perspectives from three national R&D centers focused on game-based learning: Issues in learning, instruction, assessment, and game design*. Poster presnted at the annual meeting of the American Educational Research Association, Denver, CO.

Dinn, F.S., & Calao, J. (2001). The effects of playing educational video games on kindergarten achievement, *Child Study Journal*, *31*, 95-102.

Farrar, D.E., & Glauber, R.R. (1967). Multicollinearity in regression analysis: The problem revisited, *The Review of Economics and Statistics*, *49*(1), 92-107.

Gee, J.P. (2004). Learning by design: Games as learning machines, *Interactive Educational Multimedia*, *8*, 15-23.

Huang, Z. (1998).Extensions to the k-means algorithm for clustering large datasets with categorical variables, *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 systematics: Panachea or Pandora's box?, *Annual Review of Ecology and Systematics*, *21*, 129-166.

Ke, F. (2008). A case study of computer gaming for math: Engaged learning from gameplay?, *Computers & Education*, *51*, 1609-1620.

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation, *Computers & Education*, *51*, 427-443.

Kerr, D., Chung, G.K.W.K., & Iseli, M.R. (2011). *The feasibility of using cluster analysis to examine log data from educational video games*. (CRESST Report 790). Los Angeles, CA, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Laffey, J.M., Espinosa, I., Moore, J., & Lodree, A. (2003). Supporting learning and behavior of at-risk young children: Computers in urban education, *Journal of Research on Technology in Education*, *35*, 423-440.

Merceron, A., & Yacef, K. (2004). Mining student data captured from a web-based tutoring tool: Initial exploration and results, *Journal of Interactive Learning Research*, *15*, 319-346.

Philpot,T.A., Hall, R.H., Hubing, N., & Flori, R.E.(2005). Using games to teach statistics calculation procedures: Application and assessment, *Computer Applications in Engineering Education*, *13*(3), 222-232.

R: A Language and Environment for Statistical Computing [Computer software]. (2010). Vienna, Austria: R Development Core Team.

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.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005, *Expert Systems with Applications*, *35*, 135-146.

Serrano, E.L., & Anderson, J.E. (2004). The evaluation of Food Pyramid Games: A bilingual computer nutrition education program for Latino youth, *Journal of Family and Consumer Sciences Education*, *22*, 1-16.

Squire, K. (2003).Video games in education, *International Journal of Intelligent Simulations and Gaming*, *2*, 49-62.

Sung, Y.-T., Chang, K.-E., & Lee, M.-D. (2008). Designing multimedia games for young children's taxonomic concept development, *Computers & Education*, *50*, 1037-1051.

Tobias, S., Fletcher, J.D., Dai, D.Y., & Wind, A.P. (2011). Review of research on computer games, In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction*, (pp. 127-221). Charlotte, NC: Information Age Publishing, Inc.

Vendlinski, T.P., Delacruz, G.C., Buschang, R.E., Chung, G.K.W.K., & Baker, E.L. (2010). *Developing high-quality assessments that align with instructional video games*. (CRESST Report 774). Los Angeles, CA, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Virvou, M., Katsionis,G., & Manos, K. (2005). Combining software games with education: Evaluation of its educational effectiveness, *Educational Technology and Society*, *8*, 54-65.

Warren, S.J., Dondlinger, M.J., & Barab, S.J. (2008). A MUVE towards PBL writing: Effects of a digital learning environment designed to improve elementary student writing, *Journal of Research on Technology and Learning*, *41*(1), 1613-140.