CRESST REPORT 816

USING CLUSTER ANALYSIS TO EXTEND USABILITY TESTING TO INSTRUCTIONAL CONTENT

MAY, 2012

Deirdre S. Kerr
Gregory K.W.K. Chung

National Center for Research
on Evaluation, Standards, & Student Testing
UCLA | Graduate School of Education & Information Studies

Using Cluster Analysis to Extend Usability Testing to Instructional Content

CRESST Report 816

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

May 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.

To cite from this report, please use the following as your APA reference: Kerr, D. & Chung, G.K.W.K. (2012). *Using cluster analysis to extend usability testing to instructional content.* (CRESST Report 816). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

USING CLUSTER ANALYSIS TO EXTEND USABILITY TESTING TO INSTRUCTIONAL CONTENT

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

Abstract

Commercial video games undergo usability studies to determine the degree to which the player is able to learn, control, and understand the game. Usability studies allow game designers to improve their games before they are released to the public. If usability studies could be expanded to include information about the presentation of the instructional content, they could help improve educational video games. In this study, cluster analysis was used to identify usability information from the log files from an educational video game called *Save Patch*. Cluster analysis was able to pinpoint specific levels in the game that could be improved as well as identify specific components of the level design under which certain errors were likely to occur, culminating in specific recommendations to improve the game in ways likely to increase learning.

Introduction

Video games are widely seen as having important educational potential due to their unique combination of interactive entertainment and engaging cognitive demand (Ritterfeld & Webber, 2006). However, just because a game has educational content does not necessarily mean that it will be instructionally effective (Fisch, 2005). As with any other form of instruction, the design of the material has as much or more impact on its educational effectiveness than its delivery mechanism (Nichols, 2003). Despite the best efforts of their designers, educational video games are not always easily playable (Nacke, Drachen, & Gobel, 2010), nor are they always as effective as intended. However, educational video game designers may be able to benefit from the same type of studies game designers routinely use to test the playability of their games and identify areas that could be improved. Game designers test the playability of their games via usability studies which measure the degree to which a player is able to learn, control, and understand a game (Pinelle, Wong, & Stach, 2008). Educational video game designers can use these same methods if they can find methods of expanding these traditional usability metrics to include the educational aspects of the game in order to determine not only where game mechanics break down but also where instruction fails (Nacke et al., 2010).

Since educational video games are designed to challenge users and force them to develop new skills, they cannot be expected to maintain the ease of use demanded in games designed primarily for entertainment (Pinelle et al., 2008). However, this does not mean that usability issues do not need to be addressed in educational video games, as an educational game can only be an effective instructional tool if it causes players to think about the intended instructional content (Dolmans, Gijselaers, Schmidt, & Van Der Meer, 1993). Not all usability issues are undesirable, and the decision about whether or not an identified usability issue should be corrected is best left to the researchers involved (Barendregt, Bekker, Bouwhuis, & Baauw, 2006), but it is important to identify the specific shortcomings of a game if the game is to be improved (Dolmans et al., 1993). Usability studies for educational games should identify those areas of the game in which students are not thinking about the instructional content as intended due to difficulties with either content or game mechanics. Not all educational games will be equally effective, any more than all teachers or curricula are equally effective, but thoughtful tweaking of the design of an educational game, based on identified usability issues, can increase a given game's effectiveness (Fisch, 2005).

In typical usability tests, a small number of participants play the game in question and either write down their impressions or participate in a think-aloud. However, such tests are limited by their small size, the relatively small portion of the game played by the participants, and the possible interaction between the researchers and participants (Kim et al., 2008). Identifying specific parts of a game wherein the game mechanics break down or instruction fails is better done by observing the actions players take naturally while playing the game (Barendregt et al., 2006). Logging the actions taken by players in the game allows the researcher to automatically record and calculate usability metrics of interest without interacting with the participants or artificially limiting sample size or length of play (Kim et al., 2008), and examination of the log data generated by educational video games can help researchers create a more effective learning environment by pinpointing specific usability issues that can be addressed in the next iteration of the game (Romero & Ventura, 2007).

Cluster analysis is a technique that can be used to analyze log data to identify areas of a game that have potential usability issues (Velido, Castro, & Nebot, 2011). 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 et al., 2009) through the analysis of either general correlations or sequential correlations (Bonchi, et al., 2001), and it can be used to identify specific solution strategies and error patterns students use while playing the game (Kerr, Chung, & Iseli, 2011). The resulting model of student behavior can be used to pinpoint usability issues or deficiencies of instruction (Movshovitz-Hadar, Zaslavsky, & Inbar, 1987) and thereby provide valuable information about the effectiveness of each aspect of the game (Jitendra & Kameenui, 1996).

By examining areas in a game where students use the wrong strategies more often than normal or are able to succeed using other than the intended strategies, cluster analysis can help to determine the instructional effectiveness of different portions of the game (Romero & Ventura, 2007). This study seeks to use the results of a cluster analysis to generate specific improvements that can be made to one of the games we developed, called *Save Patch*.

Method

Study Design

The data used in this study come from the log files generated by an educational video game designed by CRESST and University of Southern California's Game Innovation Lab to teach the addition of fractions called *Save Patch* (Chung et al., 2010). In this game, students are required to apply concepts underlying rational number addition to help the game character Patch bounce over obstacles to reach his home. To correctly solve each level, students must place trampolines at various locations along a one- or two-dimensional grid (see Figure 1). Students then drag coils onto the trampoline to make it bouncy. The distance Patch will bounce is the sum of all coil values added to the trampoline. For instance, if a student placed two 1/3 coils on a trampoline, Patch would bounce 2/3 of a unit.

In *Save Patch*, one whole unit is always the distance between two lines, and green dots indicate the size of the fractional pieces that should be used (see Figure 1). 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).

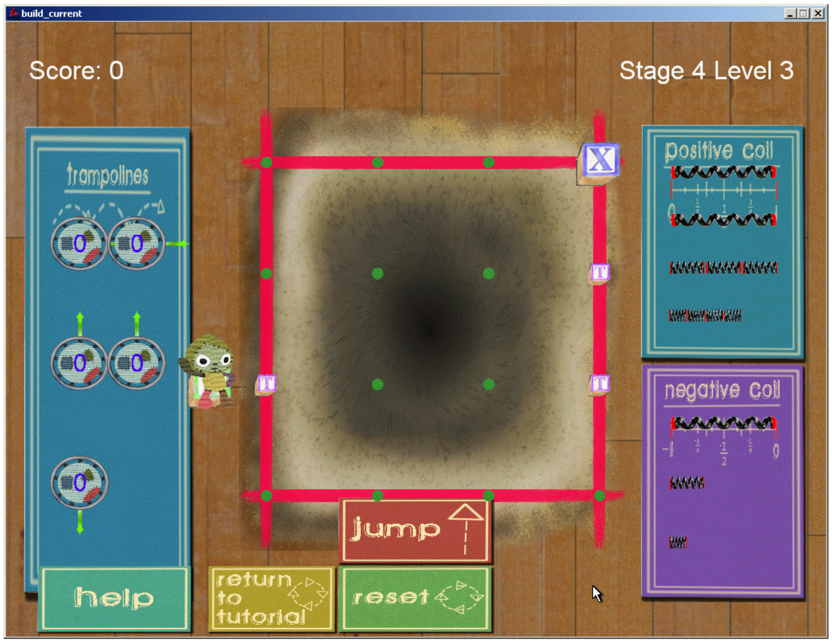


Figure 1. Screen shot of *Save Patch*.

As gameplay proceeds, trampolines must be placed at distances along the grid that are fractional parts of the whole unit. In early game levels students are given the fractional unit coils. In later levels, students are shown how to break coils into fractional units. Because only coils that have identical units can be added together, students must be attentive to what the rational number means, to what units are being added, to what units are already on the trampoline, and to how they will break coils into different size pieces. So while students could add a coil that is 1/2 a unit to another coil that is 1/2 a unit, they are not allowed to add a coil that is 1/2 a unit to a coil that is a whole unit until the whole unit is broken into two 1/2-unit coils (i.e., 2/2). When all three of these coils have been placed on the trampoline, the trampoline will show that it has 3/2 (rather than 1 1/2) units of bounce. This notation is intended to reinforce both the meaning of addition and the player's understanding of the meaning of rational numbers.

The sample of students who played *Save Patch* in this study includes 155 students (76 males and 79 females) from an urban school district in southern California. These students ranged from sixth to eighth grade and were in sixth grade math, Algebra Readiness, or Algebra 1 courses. All students played the game for approximately 40 minutes; moreover, each action the students took in the game was logged automatically.

Pattern Identification

Cluster analysis is a density estimation technique for identifying patterns within user actions reflecting differences in underlying attitudes, thought processes, or behaviors

(Berkhin, 2006). 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. 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. In the case of log data from educational video games, the identified clusters of actions reflect the different solution strategies and error patterns utilized by the students as they attempted to solve each game level (Kerr et al., 2011).

Cluster analysis on *Save Patch* identified solution strategies and six different error patterns: unitizing errors, partitioning exclusively, partitioning inclusively, misusing resources, using all resources in order, and seeing the solution as a mixed number (Kerr et al., 2011). Students who made unitizing errors failed to pay attention to the red lines that indicated the length of a unit. Instead, such students appeared to see the entire grid as one unit. Students who made partitioning errors involving counting points *exclusively* appeared to be counting the points on the grid, rather than the spaces between the points, to determine the denominator of the fraction. Students who made partitioning errors involving counting points inclusively apparently made the same error that students who counted points exclusively did, except that they also counted the points on the corners where the red lines intersected. Students who misused resources used the coils they were given in a manner that was technically correct but resulted in them running out of coils of the necessary size farther on in the level. Generally this involved using fractional units instead of a whole unit on a one-unit jump. Students who used all resources in the order in which they were provided used the order of the coils to determine which fractions to place on which trampolines, rather than examining the grid to determine mathematically which fractions to place on which trampolines. Students who saw the solution as a mixed number tried to add a whole unit and a fractional unit without first converting the whole unit to the same denominator as the fractional unit.

In order to use the cluster analysis results to check for usability issues, we calculated the percentage of attempts falling in each cluster in each level. We did this separately for the final attempt each student made in the level and the first attempt each student made in the level. The final attempt each student made at each level was examined in order to determine how each student solved the level. This was expected to be a solution except for when students failed to complete the level. If a specific error pattern accounted for a large percentage of final attempts in a given level, it would be an indication of potential usability issues. In addition, the first attempt each student made at each level was examined to determine which strategies were most common in which levels. We chose to examine first 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. Large percentages of first attempts falling in a specific error pattern may be indicative of potential usability issues, particularly if the percentage is much higher in a given level than in other levels in the game.

Results

Final Attempts

In general, one would expect the majority of final attempts at any given level to be solutions. While the percentage of attempts identified as being a solution may decrease as students reach higher levels and fewer of them manage to complete the level, the overall percentage of final attempts coded as solutions should remain fairly high and should slowly decrease as the level increases. As can be seen in Appendix A, this was the case for a majority of the levels in our game.

However, in Level 3-2 more final attempts were errors involving partitioning exclusively (46%) than were solutions (37%), indicating a potential problem with the level. As previously explained, students who made errors involving partitioning exclusively counted the points on the grid, rather than the spaces between the points, to determine the denominator of the fraction.

Closer examination of Level 3-2 (see Figure 2) revealed that the level was poorly designed. An oversight on our part made it possible for students who made errors involving partitioning exclusively to solve the level without using the desired denominator. Those students would have seen the level as requiring the use of halves instead of thirds (since they were counting dots to determine the denominator, rather than counting jumps). This would have led students to place 2/2 on the first trampoline and 1/2 on the second trampoline. When they jumped, the first jump would have taken students all the way to the end of the level, skipping over the second trampoline entirely. Since more than half of the students solved the level this way, we would recommend removing this level from the game or changing the level so that this is no longer possible. Additionally, we recommend that none of the one-dimensional levels end on a whole unit, since ending on a whole unit allows students to solve the level without using the desired denominator.

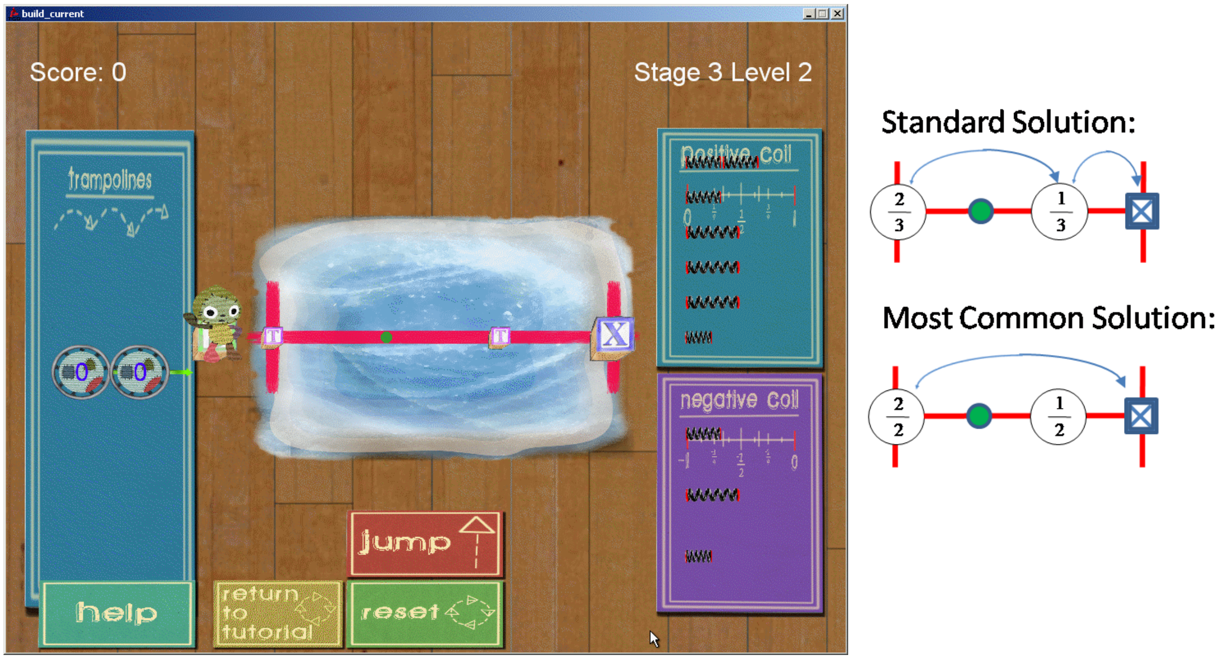


Figure 2. Screen shot of Level 3-2 of *Save Patch*, showing the most common solution to the level.

First Attempts That Were Solutions

While final attempts give us an indication of how students solved a level, first attempts give us an idea of the strategies students used. A level wherein most students' first attempt was a solution is likely a level representing material that the students already know, whereas a level in which most students' first attempts are errors is likely a level covering content students have not yet mastered. Additionally, levels having a high percentage of first attempts falling in a given error pattern may have design features that make that particular error more likely than it otherwise would be. The percentage of first attempts for each level falling in each error pattern or solution strategy are listed in Appendix B.

The initial levels of an educational game are often designed to teach the game mechanics, rather than the subject area content, so it is not surprising that a large percentage of first attempts in Level 1-1 (96%) and Level 2-1 (64%) were solutions. After that, the percentages fluctuate, with Level 4-4 (8%), Level 5-1 (6%), and Level 6-1 (6%) being the hardest. However, the percentage of first attempts that were solutions in Level 6-3 is relatively high (45%), particularly considering its late placement in the game. This may indicate that Level 6-3 is too easy for where it appears in the game.

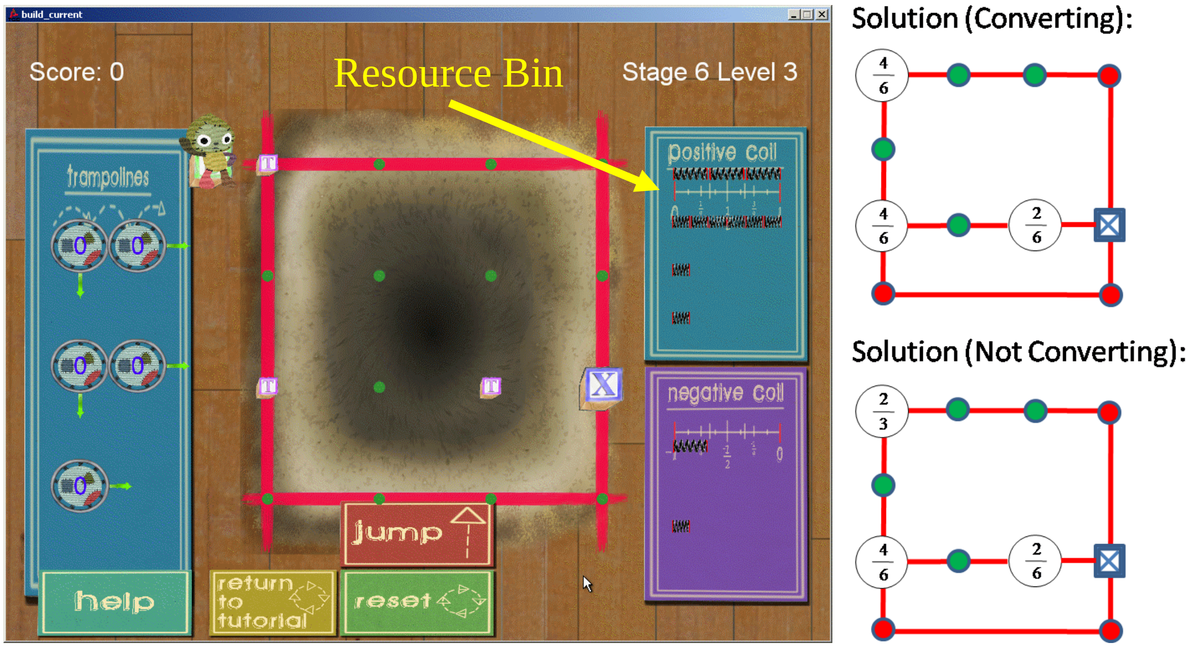


Figure 3. Screen shot of Level 6-3 of *Save Patch*, showing how students can solve the level without converting

Level 6-3 (see Figure 3) comes at the end of Stage 6, wherein students are asked to convert from the fractional units they are given to the fractional units required to solve the level. In Level 6-3 this was supposed to involve the students realizing that they did not have enough 1/6 units to solve the level, and then converting some of the 1/3 unit coils to 1/6 unit coils. However, students do not actually have to know how to convert from 1/3 units to 1/6 units, because the 1/6 unit coils are directly beneath the 1/3 unit coils in the resource bin so students can visually see that 2/6 is the same size as 1/3 without actually knowing the math. Additionally, students may not have had to convert at all, as they could have solved the level either by using 1/3 unit coils on the first trampoline and 1/6 unit coils on the second and third trampolines or by using 1/6 unit coils on the first trampoline and 1/3 unit coils on the second and third trampolines.

Since Level 6-3 can be solved without converting fractions and also has visual clues to help students convert if they choose to do so, this level may be more of a scaffolding level than a mastery level. As such, it is our recommendation that the level appear earlier in the game, perhaps as the first level in Stage 6.

First Attempts That Used All Resources In Order

Unfortunately, the most common error students made in the game was not a mathematical error. Rather, it involved students putting the coils on the trampolines in the order in which they were provided (i.e., the first row of coils on the first trampoline, the second row of coils on the second trampoline, and so on until the student runs out of trampolines or rows of coils). In Figure 3 this would have involved placing 3/3 (the first row of coils in the resource bin) on the first trampoline, 6/6 (the second row of coils) on the second trampoline, and 1/6 (the third row of coils) on the third trampoline. Using everything in order accounted for the highest percentage of first attempts in Level 3-1 (54%), Level 3-3 (37%), and Level 4-1 (39%), accounted for more than 10% of first attempts in seven other levels, and appeared in all but two levels in the game.

This error pattern is undesirable from a learning standpoint, since students can avoid learning the math by using this strategy. Because it occurs so frequently, we believe that this error may be masking mathematical errors that students would otherwise be making. Since we would like to pinpoint and remediate the mathematical errors, in later versions of the game we plan to remove the possibility of using all of the resources in order by only giving students whole unit coils on early levels in the game so students have to come up with other strategies to solve the levels.

First Attempts That Were Unitizing Errors

The first mathematical error students made was unitizing. Students who made unitizing errors failed to pay attention to the red lines that indicated the length of a unit. Instead, these students appeared to see the entire grid as one unit. This made them unable to correctly identify the denominator of the fractions required to solve the level. Unitizing errors were the highest percentage of first attempts in Level 4-5 (48%) and Level 6-1 (40%) and also occurred in Level 2-1 (15%), Level 4-1 (10%), and Level 5-1 (3%). However, there were three other levels where it was possible to make unitizing errors (i.e., where the grid size was larger than one) and none of the first attempts at these levels were unitizing errors.

It appears that students made unitizing errors most often in levels where the grid size was two whole units. The only level where students made unitizing errors that was not two whole units across was level 5-1, where only 3% of first attempts were unitizing errors. In this level, and in other levels where unitizing was possible but did not occur, the grid size was larger than one complete unit but smaller than two complete units.

An example of the type of level that encourages unitizing errors is Level 6-1 (see Figure 4) wherein 40% of first attempts were unitizing errors. The level is two complete units across and the denominator is in halves. This combination of features appears to make it easy for students to ignore the red line in the center of the grid, leading students to try to solve the level in fourths rather than in halves. Instead of than changing game levels so that this error is not possible, we recommend deliberately trapping for this error and providing feedback when it occurs to point out the unitizing error and show students how to calculate the denominator correctly.

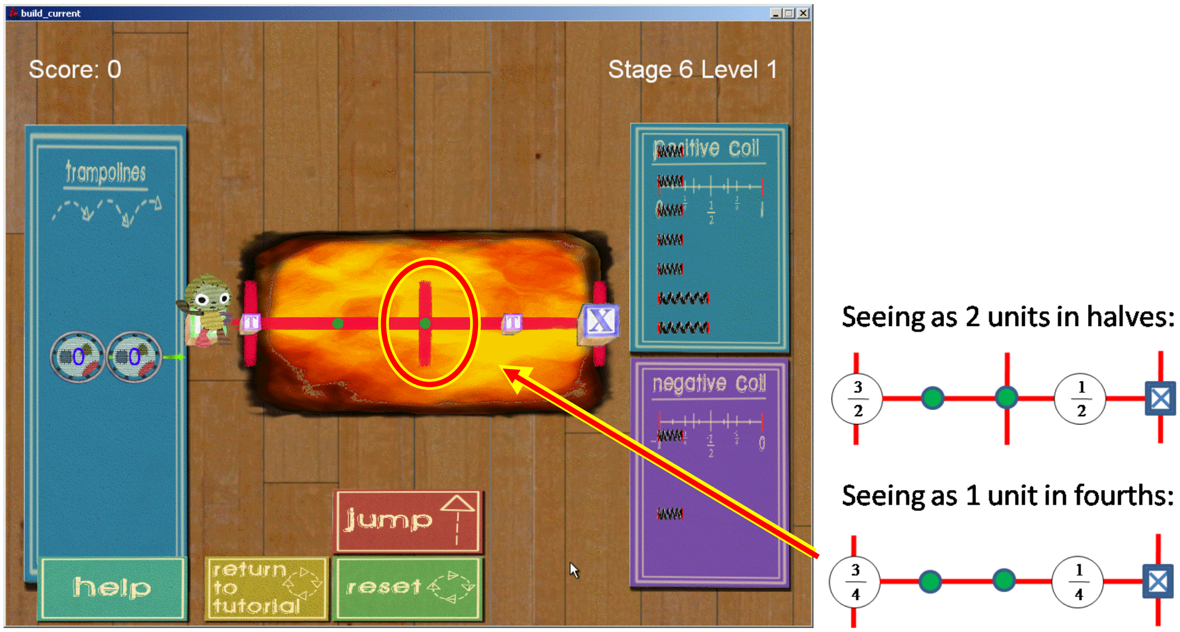
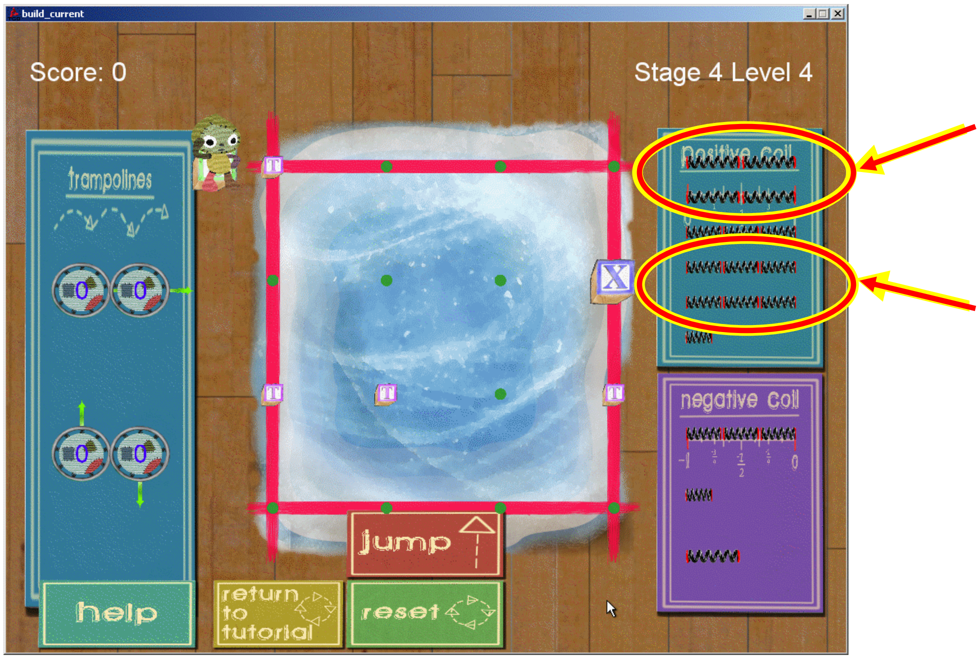


Figure 4. Screen shot of Level 6-1 of *Save Patch*, showing why students might be tempted to make unitizing errors.

First Attempts That Were Errors Involving Partitioning Exclusively

The most common mathematical error students made involved partitioning exclusively. Students who made partitioning errors determined the denominator by counting the dots between red lines rather than counting the spaces. This made them unable to correctly identify the denominator of the fractions required to solve the level. Partitioning errors were the highest percentage of first attempts in Level 3-2 (47%), Level 4-4 (49%), and Level 5-2 (49%) and accounted for more than a quarter of first attempts in Level 2-2 (26%) and Level 5-3 (25%). However, there were five other levels where it was possible to make partitioning errors (i.e., where the grid size was broken into fractional units other than halves, since a grid broken into halves would only have one dot and students are unlikely to think of one as a denominator when they are given another option) but none of the first attempts at these levels were partitioning errors.

It appears that students only make partitioning errors in levels where they have been provided fractional pieces in the resource bin that correspond to what the denominator would be if they counted dots or if they have not been given any fractional pieces and have to make the fractional pieces they desire from whole unit coils. If they are given fractional pieces in the resource bin that correspond to the correct denominator but are not given fractional pieces that correspond to counting dots, they do not make partitioning errors.



1/2 unit coils placed prominently at the top of the coil bin

1/3 unit coils placed below the 1/2 unit coils

Figure 5. Screen shot of Level 4-4 of *Save Patch*, showing why students might make partitioning errors.

An example of the type of level that encourages partitioning errors is Level 4-4 (see Figure 5) wherein 49% of first attempts were partitioning errors. The resource bin includes both third unit coils (the correct denominator for the level) and half unit coils (the denominator a student would choose if they made a partitioning error). Additionally, the half unit coils are above the third unit coils in the resource bin. This combination of features appears to make it easy for students to count the dots to determine the denominator rather than counting the spaces, leading students to try to solve the level in halves rather than in thirds. To further complicate things, as discussed earlier in this paper, students were actually able to solve Level 3-2 using partitioning errors. It is possible that without this early success, this error would not have been so prevalent (which is yet another reason to drop Level 3-2 from the game). However, rather than changing game levels so that this error is never possible, we recommend deliberately trapping for this error and providing feedback when it occurs to point out the partitioning error and show students how to calculate the denominator correctly.

Discussion

Using cluster analysis to identify the different strategies students were using to solve levels in the game allowed us to expand the traditional usability metrics to include the educational aspects of the game in order to determine where instruction failed. Specific deficiencies in instruction were pinpointed by examining final attempts to determine where students were able to succeed using other than the intended strategies and examining first attempts to determine where students were most prone to using erroneous strategies. This gave us the ability to suggest specific changes to the game that could increase learning and address specific misunderstandings.

In *Save Patch*, examining the final attempts students made at each level to determine the percentage of students who solved that level using each of the strategies identified by the cluster analysis allowed us to discover usability issues of which we were previously unaware. This analysis pinpointed a specific level which students were able to accidentally solve using a common mathematic error pattern. The game would likely benefit from removal of this level, as it seems feasible that students may be internalizing incorrect assumptions about the math if it is possible to solve levels erroneously.

In addition, examining the first attempts students made at each level to determine the percentage of students using each of the strategies in their initial approach allowed for the discovery of additional usability issues of which we had been unaware. This analysis revealed that students were frequently avoiding doing the math by adopting an “everything in order” strategy. For this reason it might be better to provide the students with only whole unit coils on early levels of the game, rather than fractional unit coils, so that students will have to choose the coil size they desire rather than simply placing all the coils on the trampolines in the order in which they were provided in the resource bin.

Analysis of first attempts also allowed for the identification of level features that appear to make certain mathematical errors more likely. If the grid size is exactly two units across, students tend to make unitizing errors. If the resource bin includes coils corresponding to what the denominator would be if students counted dots instead of spaces to determine the denominator, students make partitioning errors. Later in the game students also make partitioning errors when given only whole unit coils, but do not make these errors when given coils with the correct denominator (and not coils with the denominator that would result from counting dots).

Since this method resulted in the identification of specific improvements that could be made to potentially increase the game’s effectiveness, it appears that cluster analysis can be used to extend usability testing to instructional content. This may enable researchers to design a more effective instructional parcel, as it identifies areas in the game in which students are not thinking about the material as intended, which may in turn help educational games begin to live up to their potential as instructional instruments.

References

Barendregt, W., Bekker, M. M., Bouwhuis, D. G., & Baauw, E. (2006). Identifying usability and fun problems in a computer game during first use and after some practice. *International Journal of Human-Computer Studies*, *64*, 830-846.

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., Pedreschi, D., ... Ruggieri, S. (2001). Web log data warehouses and mining for intelligent web caching. *Data & Knowledge Engineering, 39,* 165-189. doi:10.1016/S0169-023X(01)00038-6

Chung, G. K. W. K., Baker, E. L., Vendlinski, T. P., Buschang, R. E., Delacruz, G. C., Michiuye, J. K., & 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.

Dolmans, D. H. J. M., Gijselaers, W. H., Schmidt, H. G., & Van Der Meer, S. B. (1993). Problem effectiveness in a course using problem-based learning. *Academic Medicine, 68,* 207-213.

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

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. doi:10.1023/A:1009769707641

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. doi:10.1146/annurev.es.21.110190.001021

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

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.

Kim, J. H., Gunn, D. V., Schuh, E., Phillips, B. C., Pagulayan, R. J., & Wixon, D. (2008). Tracking real-time user experience (TRUE): A comprehensive instrumentation solution for complex systems. In *Proceedings of the 26th annual SIGCHI Conference on Human Factors in Computing Systems,* 443-452.

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,* 3-14.

Nacke, L., Drachen, A., & Gobel, S. (2010). Methods for evaluating gameplay experience in a serious gaming context. *International Journal of Computer Science in Sport*, *9*(2), 1-12.

Nichols, M. (2003). A theory for eLearning. *Educational Technology*, *6*(2), 1-10.

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

Ritterfeld, U., & Weber, R. (2006). Video games for entertainment and education. In P. Vorderer & J. Bryant (Eds.), *Playing video games: Motives, responses, and consequences* (pp. 388-413). Mahwah, NJ: Erlbaum.

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. doi:10.1016/j.eswa.2007.11.026

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

Velido, A., Castro, F., & Nebot, A. (2011). Clustering educational data. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S.J.d. Baker (Eds.), *Handbook of educational data mining*. Boca Raton, FL: CRC Press.

Appendix A:
Final Attempt Distribution

Table A1 shows the percentage of final attempts for each level that fell in each cluster. UT stands for unitizing errors, PE stands for errors involving partitioning exclusively, PI stands for errors involving partitioning inclusively, MR stands for misusing resources, IO stands for using all resources in the order in which they were provided, and MN stands for seeing the solution as a mixed number.

Table A1

Percentage of Final Attempts in Each Cluster

		Error patterns					
Stage - Level	Solutions	UT	PE	PI	MR	IO	MN
1-1	99%	--	--	--	--	--	--
2-1	94%	0%	--	--	2%	3%	--
2-2	100%	--	0%	--	0%	0%	--
3-1	94%	--	--	--	--	3%	--
3-2	37%	--	46%	--	--	11%	--
3-3	99%	--	--	--	0%	1%	--
4-1	97%	1%	--	--	--	1%	--
4-2	92%	--	--	0%	2%	1%	--
4-3	83%	--	--	--	0%	3%	--
4-4	87%	--	4%	0%	--	3%	--
4-5	95%	1%	--	--	1%	0%	--
5-1	80%	1%	--	--	--	0%	1%
5-2	78%	--	3%	0%	--	0%	--
5-3	77%	--	3%	1%	--	0%	--
6-1	77%	3%	--	--	2%	4%	--
6-2	56%	--	--	--	--	4%	--
6-3	74%	--	--	0%	--	6%	--

Appendix B:
First Attempt Distribution

Table B1 shows the percentage of first attempts for each level that fell in each cluster. UT stands for unitizing errors, PE stands for errors involving partitioning exclusively, PI stands for errors involving partitioning inclusively, MR stands for misusing resources, IO stands for using all resources in the order in which they were provided, MN stands for seeing the solution as a mixed number, and UE stands for unknown error. X's indicate levels in which a cluster was possible, but was not committed by students.

Table B1

Percent of First Attempts in Each Cluster

		Error patterns						
Stage - Level	Solutions	UT	PE	PI	MR	IO	MN	UE
1-1	96%	--	--	--	--	--	--	--
2-1	64%	15%	--	--	2%	8%	--	8%
2-2	24%	X	26%	X	1%	X	--	40%
3-1	41%	--	--	--	--	54%	--	5%
3-2	16%	--	47%	X	--	12%	--	21%
3-3	32%	X	X	X	9%	37%	--	14%
4-1	31%	10%	--	--	X	39%	--	10%
4-2	44%	X	X	18%	3%	2%	--	26%
4-3	39%	--	X	X	3%	17%	--	24%
4-4	8%	--	49%	2%	--	13%	--	32%
4-5	17%	48%	--	--	1%	6%	--	12%
5-1	6%	3%	--	--	--	15%	30%	39%
5-2	13%	--	49%	0%	--	23%	--	31%
5-3	16%	--	25%	4%	--	4%	--	55%
6-1	6%	40%	--	--	18%	7%	X	35%
6-2	31%	--	X	X	--	17%	--	41%
6-3	45%	--	X	5%	--	22%	--	23%