CRESST REPORT 808

TEACHING RATIONAL NUMBER ADDITION USING VIDEO GAMES: THE EFFECTS OF INSTRUCTIONAL VARIATION

NOVEMBER, 2011

Terry P. Vendlinski
Greg K. W. K. Chung
Kevin R. Binning
Rebecca E. Buschang

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

Teaching Rational Number Addition Using Video Games:
The Effects of Instructional Variation

CRESST Report 808

Terry P. Vendlinski, Gregory K. W. K. Chung, Kevin R. Binning, and Rebecca E. Buschang
CRESST/University of California, Los Angeles

November, 2011

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: Vendlinski, T.P., Chung, G.K.W.K., Binning, K.R., & Buschang, R.E. (2011). *Teaching rational number addition using video games: The effects of instructional variation*. (CRESST Report 808). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

TABLE OF CONTENTS

TEACHING RATIONAL NUMBER ADDITION USING VIDEO GAMES: THE EFFECTS OF INSTRUCTIONAL VARIATION

Terry P. Vendlinski, Gregory K. W. K. Chung, Kevin R. Binning, and Rebecca E. Buschang
CRESST/University of California, Los Angeles

Abstract

Understanding the meaning of rational numbers and how to perform mathematical operations with those numbers seems to be a perennial problem in the United States for both adults and children. Based on previous work, we hypothesized that giving students more time to practice using rational numbers in an environment that enticed them to apply their understanding might prove educationally beneficial. We developed a video game, based on two key ideas about addition and rational numbers, to investigate this hypothesis. We also analyzed the effects of different types of feedback provided to students during the videogame. Our findings in this initial study suggest that designing such a video game is not only possible, but also that students using a game designed in this manner can increase their ability to add rational numbers even when playing the game for a relatively short period of time. Since the effect size of a single 40-minute intervention is moderate, we discuss the need for future studies designed to spread game play over several class periods and to include instructional resources external to the game. We discuss implications for the larger efficacy study to follow.

Introduction

Students (and many adults) in the United States continue to have difficulty understanding the meaning of rational numbers and how to perform mathematical operations with those numbers despite numerous attempts to address such shortcomings (Misquitta, 2011; NCTM, 2000; Siebert & Gaskin, 2006; U.S. Department of Education, 2008). While many efforts to remediate these deficits have been made, few have succeeded (see for example, Beesley, Apthorp, Clark, Wang, Cicchinelli, & Williams, 2011; Garet et al., 2011). Programs that have been successful have often focused on getting students and teachers to understand how key foundational ideas in a domain like rational numbers relate to one another and how these ideas are applied to solve seemingly dissimilar problems (e.g., Carpenter, Fennema, Franke, Levi, & Empson, 2000; Phelan, Choi, Vendlinski, Baker, & Herman, in press). We recently completed the study of such a program, designed in part, to help teachers understand how to teach rational number concepts to middle school students (Vendlinski & Phelan, 2011; Phelan et al., 2011). Based on those successes and on other experiences, we hypothesized that giving students more opportunities to actually apply those key foundational ideas would improve their understanding of and ability to apply rational number concepts in problem solving.

Given the popularity of video games (Flew & Humphreys, 2005) and the large amount of time young Americans spend playing them (Kaiser Family Foundation, 2002), many have wondered whether designing instruction into video games might help students learn better or learn more (Gee, 2003). The results of past educational interventions using video games seem mixed (Kebritchi, Hirumi, & Bai, 2010). In fact, recent research has suggested that the belief that video games will intrinsically motivate students to learn may be erroneous (Charsky & Ressler, 2011; Hamlen, 2011); however, we speculated that designing video games around a limited number of key foundational concepts and inviting students to play the game by applying those concepts would prove beneficial to learning. Our prior research suggested that such an approach should be studied. We also wanted to study the effects of different types and formats of feedback during in-game instruction. Our research questions were as follows:

1) Can a video game be designed that helps students learn important mathematical concepts using minimal classroom time?
2) Do different treatments of video game instruction or feedback produce different effects on student learning?
3) Is a one class period interaction with the game adequate to produce average student outcomes on the posttest that are commonly viewed as acceptable (i.e., greater than 70% correct)?
4) Do different treatments of video game instruction or feedback produce differential effects for different types of students?
5) What other research questions should be answered prior to the full efficacy study?

In this report, we describe a study that was designed to inform a future efficacy study. We tested the effects of several video game interventions to estimate the effect sizes associated with these interventions and to determine which, if any, might be most promising for the subsequent efficacy study. Consequently, we skewed the size of various treatments in favor of the interventions that had previously shown promise or that the literature suggested might produce larger effects than interventions we had previously tested. A small number of students in each class were also assigned to a control condition in which they played a math video game unrelated to rational number addition. We describe the most promising and statistically significant of these effects, the version of the game that seems most generally useful, as well as when alternative game instantiations might be warranted.

Methods

The Sample

Two California school districts agreed to participate in the field study described in this paper. In the first district, the participants were all suburban 6^{th}, 7^{th}, and 8^{th} grade middle school students in Southern California. These students were either enrolled in sixth-grade math, in an Introduction to Algebra course, or in Algebra 1. The second district was a rural district in California's San Joaquin Valley. Ninth graders in this district were enrolled in either pre-algebra or first year algebra. The tenth, eleventh, and twelfth graders from this district who were involved in this study were all enrolled in first year algebra. In addition, this district also enrolled some of their algebra students in a two-period math course where students studied algebra in the first period and prepared to take the California High School Exit Exam (CAHSEE) or participated in a period of extended algebra study during the second period. These courses were termed Algebra Success/CAHSEE or Algebra Success/Algebra, respectively. The 365 subjects involved in this study represent a sample of convenience drawn from in situ math classrooms. Table 1 shows the number of students in each course, by grade.

Each district established their own policies for assigning students to classes. Aside from students in the sixth grade, who were all in sixth grade math, scores on a student's previous California Standards Test (CST) and previous math teacher recommendation were the primary basis for class assignment in both districts. In the middle school district, most eighth-grade students took first year algebra unless their seventh grade teacher felt they were not ready for that course. Those eighth graders not assigned to an algebra class, in this district, were assigned to the Introduction to Algebra course. Most seventh graders in our sample were enrolled in Algebra 1.

Table 1

Sample Size by Grade Level Within Each Math Class Type

Grade level	Number of subjects
Algebra 1	
7th	16
8th	29
9th	87
10th	54
11th	10
12th	1
Unknown	9
Pre-algebra	
9th	47
Unknown	1
Sixth grade math	
6th	25
Introduction to Algebra	
7th	2
8th	17
Unknown	1
Algebra Success/CAHSEE	
9th	24
Unknown	1
Algebra Success/Algebra	
9th	38
Unknown	3

In the high school district, many of the incoming ninth graders also took first-year algebra. Students in ninth grade who were not determined to have the necessary prerequisite knowledge or math skills were assigned pre-algebra. To matriculate from high school in this district, every student was required to pass two years of high school math—one year of which had to include algebra. Students enrolled in either of the Algebra Success classes were only counted as part of that class, and not the Algebra 1 class, in this study.

With the exception of the Algebra 1 class, the grade level of students in each of the other classes was largely homogenous. The heterogeneity of grade level in Algebra 1 is, in part, attributable to districts moving toward California's stated goal for all eighth grade students to take algebra (California State Board of Education, 1997) and to a 2007 decision to allow seventh graders to take Algebra 1. Although that goal has now been modified with the adoption of the Common Core Mathematics Content Standards (California State Board of Education, 2010), approximately half of California's eighth graders take Algebra 1, and a substantial number of high school students still take Algebra 1 either because they must repeat the class or because they were not offered the class as eighth graders. In addition to taking Algebra 1 later, these high school students differ from the middle school algebra students in that the likelihood of passing the CST for Algebra 1, as either a repeat or as a first-time test taker, decreases substantially after eighth grade (Vendlinski, 2011).

The *Save Patch* Rational Number Addition Game

The students in this study were divided into six groups. Five of the six groups played some version of a video game that involved rational number addition and one group (the control) played a video game that focused on using mathematical operations to rewrite mathematical expressions.

In the rational number addition video game (called *Save Patch*), students were presented with the challenge of bouncing a small sack-like doll (Patch) over various hazards in order to get it safely to the other side of the hazard. To do so, students were asked to place trampolines at various fixed locations along a one- or two-dimensional grid. Students made each trampoline "bouncy" by dragging coils onto the trampoline. The distance each coil caused Patch to bounce was commensurate with its length and the grid. Therefore, if a student added a coil of one unit to a trampoline, that trampoline caused Patch to bounce exactly one unit on the grid. A screen shot of the game in shown in Appendix A.

Students in all treatment conditions learned that in *Save Patch,* one whole unit was always the distance between two red lines. It was this unit that became the referent for coils of fractional bounce later on. Coils could be added to a trampoline to increase the distance Patch would bounce; however, only identical coils could be added together (whole coils to whole coils, thirds to thirds, fourths to fourths, etc.). While students could place any size coil on the trampoline initially, subsequent coils could only be added to the trampoline if they were the same size. Initially, students were asked to add whole unit (integer) coils to a trampoline one at a time, to reinforce the *meaning* of addition with integers. While the game had an option to include negative coils, this feature was not used in this study.

The *Save Patch* game exploits the fact that real numbers can be broken into smaller, identical parts (decomposed), if necessary, to facilitate addition and that this process is similar in both integer and rational number (fractional) addition. The intent is to make explicit connections between integer addition (with which many students have confidence) and fractional addition (with which many students struggle). Moreover, the game play requires that players (students) be attentive to the size of a unit they are adding. Fluency with these basic ideas is integral, not ancillary, to game play (i.e., the game mechanic) in *Save Patch*.

As game play proceeded, students were required to place trampolines at distances along the grid that were fractional parts of the whole unit. Consequently, students were first given and then shown how to break coils into proper fractional units. Since only identical units could be added together, students had to be attentive to what the rational number meant, to what units were being added, to what units were already on the trampoline, and to how they would break the given coils into different sized pieces. This game feature was intended to reinforce both the meaning of addition and to reinforce the player's understanding of the meaning of rational numbers.

Since *Save Patch* was focused on the addition of rational numbers, the conversion of fractions of different sizes (i.e., fractions with different denominators) was not accomplished through multiplication. In fact, the understanding of that process was beyond the specified learning goals (knowledge specifications) around which the *Save Patch* game was designed. Rather, students were shown how they could use the mouse to click on a coil and then scroll up or down to break the coil into more pieces (each smaller in size) or fewer pieces (each larger in size), respectively.

The standard symbolic representation of a fraction ($\frac{\#}{\#}$) was shown alongside each coil as the student scrolled on the coil. For example, if a student clicked on a coil that was one whole unit in length and scrolled up, the coil broke first into two halves, then three thirds as the student scrolled up again, etc. If the student used the same procedure after clicking on a 1/2 coil, then the coil broke into two fourths, and scrolling again would produce three sixths, etc. As long as students did not click somewhere else on the game, they could also scroll down on these same coils to make fewer pieces that were larger in size (for example, the student could scroll three sixths to make two fourths or one half).

As shown in the Appendix A, the grid representation was also used to convey the meaning and use of rational numbers. As mentioned previously, one whole unit was always the space between two red lines. In the one-dimensional game, the red lines denoting *unit* were vertical, and in the two-dimensional game these *unit* lines were both vertical (counting units across the screen) and horizontal (counting units up the screen). Fractional parts of that unit distance were represented as the distance between green dots placed equidistant between red lines along the grid. At times, trampoline blocks were placed over the green dots, so students quickly learned or, in some versions of the game, were told that trampolines or blocks between solid red lines also meant that the whole had been divided into smaller pieces.

Prior to the present study, we had tested the game with various amounts and forms of onscreen textual instruction and feedback (Chung et al., 2011; Delacruz, 2011), but based on the work of Mayer and others (Baddeley, 1999; Mayer, 2005; Sweller, 1999), we suspected that video-based instruction and feedback might be more effective than text-based instruction and feedback for both English language learners as well as for those proficient in English. In this study, therefore, we included these types of feedback as additional conditions.

In all, five treatment versions of *Save Patch* were developed to test the impact of tutorial and feedback variations on math and game outcomes. We called the game with mechanics only instruction the *baseline* condition. By way of a graphics-based primer on the game mechanic, this condition only informed players of the goal of the game and the tools available to the player so they could achieve the goal. For example, this game condition taught students how to drag coils onto the trampolines, how to move the trampolines onto the grid, and how to scroll. Mathematical references in the baseline condition were minimized as much as possible—as the instruction was intended to teach students how to play the game rather than to increase understanding of how a unit was defined, rational numbers and their relationship to that unit, or addition. Graphics included both text and images of the game screens.

The second version of the game (graphics-based mechanics instruction and video feedback) also gave students a graphics-based primer on the game mechanic, but this version of the game also monitored an individual student's game play and provided video-based feedback to the student when it detected that a student had made incorrect moves. The game was programmed to intervene with feedback after a selected number of errors. In this case, the game ignored the first error, but if the error continued, the game alerted the player to a specific error after two errors, and suggested that the player focus on a certain misconception (e.g., counting breaks rather than spaces to determine the denominator, etc.). If the student continued making the same error, the game would deliver video-based feedback showing the student what specific actions to take to resolve the error.

A third treatment condition (graphics-based math instruction and feedback) provided students graphics-based instruction on how to play the game. In addition to the basic game mechanics instruction that students in the first two conditions received, this treatment incorporated specific math instruction. In particular, the instruction focused on how the unit was used to define a fraction, how to use the number of pieces a unit was broken into to define the denominator of a fraction, how to determine the value of the numerator to determine the number of equal size pieces needed to jump a particular distance, and how addition of equally sized pieces might be used in the game. In addition, this condition also provided graphics-based feedback to the student after a selected number of errors. As in the condition above, the first error was ignored, but if the error was made again, the game alerted the player to the error. If the error persisted, the player was alerted to their specific mistake and, eventually, shown how to resolve the error. In this condition, however, both instruction and feedback were graphics based.

The fourth and fifth treatment conditions were variants of the third treatment condition. In both cases, students received instruction before playing certain levels of the game, and students were also provided feedback in the game when they made mistakes. Unlike the instruction in the previous condition, however, the instruction provided to students in the fourth and fifth treatment groups was all video-based instruction. As before, the instruction focused on how the unit was used to define a fraction, how to use the number of pieces a unit was broken into to define the denominator of a fraction, how to determine the value of the numerator to determine the number of equally sized pieces needed to jump a particular distance, and how addition of equally sized pieces might be used in the game. The only difference between each of these two conditions was in how the game delivered the feedback. In the fourth condition (video-based math instruction with graphic-based feedback), the initial instruction was delivered using video, but feedback was provided using graphics. In the fifth condition (video-based math instruction and feedback), the student player received all the instruction and feedback in a video-based format.

In the control condition, the students played a video game designed to teach the meaning of the operations of addition, subtraction, multiplication, and division and the effects of these operations on expressions. No fractions were involved in this game. The students in this condition played their game for the same amount of time and completed the same pretest as did the students who played *Save Patch*. With one exception, these students also received the same posttest as their peers in the treatment groups. The exception was that students in the control group were not asked questions on the posttest that referred to the *Save Patch* game. For example, the control students were not asked how far Patch would jump if $\frac{4}{3}$ of a coil were on the trampoline. Consequently, the findings in this study only involve the pretest and posttest items that were presented to both groups.

As indicated previously, this study was intended to be a precursor to a larger efficacy trial. Consequently, our purpose was to test a number of interventions in order to estimate the effect sizes of various interventions and to determine which, if any, interventions might be most promising for the subsequent efficacy study. Given that the samples reported in this study were small samples of convenience, we did not design a fully crossed, factorial study at this time. Rather, based on our prior experience (Vendlinski, Delacruz, Buschang, Chung, & Baker, 2010), we assigned more students to those interventions that we thought likely to produce (or reproduce) significant pre- to posttest gains after 40 minutes of game play. As a consequence, not all groups had significant statistical power to reject various hypotheses for every intervention.

To this end, more students were given the video or video and graphics-based interventions than were given the graphics only or minimal instruction interventions that we had previously evaluated. For comparison purposes, a small number of students in each class were also assigned to the control condition that played a math video game unrelated to rational number addition between the pretest and posttest.

Pretest and Posttest

Regardless of treatment condition, each student was given the same pretest prior to game play. The items on the pretest were based on a small number of knowledge specifications (learning objectives) that are given in Appendix B. The pretest was designed to test both conceptual and procedural understanding of this knowledge and had undergone extensive analysis to assure high technical quality prior to this study. The procedure used to determine the technical quality of the tests is described in detail elsewhere (Vendlinski et al., 2010).

The posttest consisted of all items that appeared on the pretest. Students who had played the *Save Patch* game were also asked several additional questions about rational number addition using the *Save Patch* game representation. Each student in the control group received an identical pretest and posttest. As stated above, the comparisons made between the treatment and control conditions in this study only involve the pretest and posttest items that were presented to both groups.

We determined the reliability of the pretest and posttest by calculating inter-item reliability on both the pretest and the posttest, and the pretest–posttest correlation between percent correct on the pretest and percent correct on the posttest for the control group. As the control group received no instruction on rational numbers, we expected significant correlations between the percent correct scores on both tests. As has been our practice when using identical items on the pretest and posttest, we also tested for significant pretest to posttest gains in the control group to ensure that students had not improved merely because they learned about rational number addition from taking the pretest.

Surveys

Each student was also given two surveys. The first survey asked students several questions about their background, including grade level, gender, and previous math grade. This survey was given in conjunction with the pretest. The second survey was given in conjunction with the posttest. This second survey asked students about their attitudes toward math, their video game play behaviors, and their thoughts about the specific game they played during the study. While the primary purpose of the second survey was to inform the full efficacy study, we did use game play behaviors (e.g., each student's self-reported amount of weekly video game play) in the various analyses reported in this study. The surveys were given in two parts so as to minimize student "test" fatigue.

Choosing an Appropriate Data Set for Analysis

Students in the study were asked to complete all items on the pretest and the posttest or to write "I don't know" (IDK) by those items they could not finish. A number of students in the study, however, left items blank on the pretest and on the posttest. We became concerned that recoding these blanks as incorrect answers might adversely affect the accuracy of our analysis. Merely recoding a missing pretest response as incorrect could underestimate the preexisting knowledge of students, while recoding such responses on the posttest could underestimate the effects of the game. On the other hand, merely dropping a student who had any missing data would seem likely to produce inaccurate estimates of the game's effectiveness since the remaining data would likely have fewer incorrect responses. Rejecting these two extreme courses of action—recoding all missing as incorrect or dropping cases with any missing data—required that some other objective method be employed to address missing data before the data set could be analyzed.

We explored two methods to determine whether a student was included or excluded from further analysis. First, we looked for natural breaks in the data that might indicate which students to exclude from the sample. Second, we looked at the randomness of the missing data for each student to decide if the student should or should not be included in further analysis. In both cases, if a student was selected for inclusion in the data set for further analysis (the reduced sample), we recoded missing responses as incorrect answers, and we then ran descriptive and crosstab analyses on key demographic characteristics to determine if the reduced student sample differed significantly from the complete sample. Previous studies suggested that certain characteristics were either highly correlated with the pretest, with the posttest or with game success (see Vendlinski et al., 2010) so we were interested in demonstrating that the complete and reduced samples were statistically similar in this regard. In particular, gender and previous year's math grade had shown high correlations with both tests and game play success, while amount of weekly game play showed high correlation with game success alone. In addition, we hypothesized that test effort, perception of test difficulty, and perception of the importance of the test might contribute to completion rates and, therefore, we wanted to assure ourselves that the samples were not dissimilar on these important characteristics.

Our first effort to cull missing responses from the data was to eliminate cases where students had not responded to more than six items on either test. We chose six items as the cut-point because we had observed that there seemed to be a substantial decrease in the number of students leaving more than six items blank compared to the number leaving fewer than six items blank on the pretest and the posttest. While a number of students left six or fewer items blank on the pretest or the posttest, substantially fewer students left more than that number of items blank on either test.

Using the natural breaks in the data, we eliminated 39 students. Unfortunately, this culling of students *did* result in the complete data set and the reduced data set being significantly different on key variables (as seen in the crosstab analyses shown in Table 2), namely on the variables of perceived test difficulty and effort to do well on the tests.

Table 2

Comparison of Key Variables in the Complete and Reduced Data Sets After Using a Cut-Score Culling Procedure

Variable	χ^2	*df*	*p*
Gender	0.295	1	.587
Weekly amount of video game play	7.876	4	.096
Ethnicity	3.074	6	.799
Previous year's math grade	7.191	4	.126
Difficulty of pretest	16.277	3	.001***
Effort made to do well on pretest	7.910	3	.048*
Student's perception that pretest was important	3.253	4	.516
Difficulty of posttest	34.244	3	< .001***
Effort made to do well on the posttest	7.090	3	.069
Student's perception that posttest was important	0.609	4	.962

*$p \leq .05$. **$p \leq .01$. ***$p \leq .001$.

Given these results, we created another reduced sample data set based on the nature of the items students left unanswered. In these efforts, we tried to discern random versus non-random patterns in student responses that would account for the large number of blank items. For example, when students left large numbers of items at the end of the test blank and also left the last items on the test blank, we concluded that these students may have run out of time to complete the test or had become fatigued and just chose not to complete the test. Therefore, we were hesitant to infer the student did not know these items and then further infer a missing response was an incorrect answer. Instead, we proposed to drop these students from further analysis. We also proposed to eliminate students who had skipped random sections of the test. We argue that these students were different from students who skipped sections of the test which asked about a specific concept such as adding fractions or representing fractions on a number line. On the other hand, students who showed a systematic avoidance of certain problem types, such as students who skipped all problems that involved addition of fractions with unlike denominators or who skipped all items asking them to represent fractions on a number line, seemed to avoid such problems because they were unable to answer them. We proposed to keep this latter group of students in our reduced data set and to recode their missing responses as incorrect answers.

By analyzing the response patterns of students with missing data in this way, we identified a total of 16 students who we judged should be dropped from further analysis. After dropping these students and recoding any remaining missing responses as incorrect answers, we again used a chi-square analysis to compare the complete data set to this reduced data set on the variables of interest described above. The results of this analysis are provided in Table 3 below.

Table 3

Comparison of Key Variables in the Complete and Reduced Data Sets Using a Response Pattern Culling Procedure

Variable	χ^2	*df*	*p*
Gender	< 0.001	1	.991
Weekly amount of video game play	0.066	4	.999
Ethnicity	0.093	6	1.000
Previous year's math grade	0.100	4	.999
Difficulty of pretest	0.184	3	.980
Effort made to do well on pretest	0.011	3	1.000
Difficulty of posttest	0.084	3	.994
Effort made to do well on the posttest	0.050	3	.997
Student's perception that pretest was important	0.036	4	1.000
Student's perception that posttest was important	0.105	4	.999

The reduced sample that resulted was statistically identical to the complete sample on the key demographic variables thought to be associated with game play, with effort to perform well on the test, and with perceived test difficulty. Moreover, the reduction allowed us to remove students who chose not to or did not have time to finish both tests, which could arguably cause inaccurate estimates of treatment effects.

Normality Assumptions of Classroom Populations

Given our intent to investigate differences in mean scores and estimate treatment effect sizes, we next tested the assumption that the pretest and posttest scores were approximately normally distributed using a one sample Kolmogorov-Smirnov (K-S) test. Neither the pretest scores ($Z = 1.935$, $p = .001$) nor the posttest scores ($Z = 2.\ 321$, $p < .001$) of our reduced sample proved to be statistically normal in their distributions. We suspected that this might be an artifact of the sample of convenience and the fact that two disparate districts were being combined.

Consequently, we next evaluated the normality of the pretest and posttest distributions by class type since every class except algebra was composed of students from just one district and largely from one grade level. In every case, but one, our analysis suggested that pretest and posttest scores were normally distributed within a particular type of class. The results of our analysis of both pretest and posttest distributions by class type for the full as well as the reduced sample are provided in Table 4 below. In addition, the Mann-Whitney statistic was calculated to test the hypothesis that the mean of the reduced sample was statistically equivalent to the mean of the complete sample on the pretest and on the posttest for each class type.

Table 4

Normality and Mean Equivalency Tests, by Class Type, for the Complete and the Reduced Data Sets on Both the Pretest and the Posttest

	Pretest K-S statistic						Posttest K-S statistic						Mann-Whitney test							
	Complete sample			Reduced sample			Complete sample			Reduced sample			Pretest				Posttest			
Class type	*Z*	*n*	*p*	*Z*	*n*	*p*	*Z*	*n*	*p*	*Z*	*n*	*p*	*U*	n_1	n_2	*p*	*U*	n_1	n_2	*p*
Algebra 1	1.427	206	.034[a]	1.405	202	.039[a]	1.681	205	.007[a]	1.657	202	.008[a]	20700.0	206	202	0.929	20585.0	205	202	0.919
Pre-algebra	1.303	48	.067	1.264	45	.082	1.285	48	.073	1.215	45	.104	1051.0	48	45	0.823	1070.5	48	45	0.942
Sixth grade math	0.505	25	.960	0.505	25	.960	0.644	25	.801	0.644	25	.801	312.5	25	25	1.000	312.5	25	25	1.000
Intro to Algebra	0.903	20	.389	0.796	18	.551	0.481	20	.975	0.548	18	.925	167.0	20	18	0.718	174.0	20	18	0.874
Algebra Success/ CAHSEE	0.986	25	.285	0.940	23	.341	0.747	25	.632	0.719	23	.679	278.5	25	23	0.852	285.0	25	23	0.959
Algebra Success/ Algebra	1.195	41	.115	1.072	36	.201	1.056	40	.214	0.818	36	.516	715.5	41	36	0.818	693.5	40	36	0.783

[a]The K-S statistic is significant indicating that the sample distribution is significantly different from the normal distribution.

As can be seen in Table 4, neither the pretest scores nor the posttest scores for students in the Algebra 1 class were normally distributed in either the complete or reduced sample data set. In fact, both tests generated results that were bimodal in their distribution. This was not surprising given that the students came from two disparate districts and that the students in the suburban district took algebra at or before the grade level mandated by the state of California, whereas students in the rural district (re)took algebra after that grade.

Although the bimodal distribution of the data might suggest dividing the sample into two subsamples at the mean, such a division could be problematic. Specifically, dividing the groups in this way could, by definition, create an interaction between the pretest or posttest score and the resulting class group for students scoring above the mean and students scoring below the mean. Such a grouping would also be artificial rather than reflecting the fact that the students were actually sampled from two distinct groups. With this in mind, we also investigated whether students above and students below the mean on the pretest were equivalently distributed across grade levels. A chi-square analysis χ^2 (5, n = 194) = 69.69, $p < .001$ suggests that the reclassification of students into the high and low algebra sub-groups, based on mean pretest score, is not independent of grade. The dependency of grade level and mean-based subgroups is evident in Table 5.

Table 5

Cross-Tabulation of Students Above or Below the Overall Mean Score on the Pretest by Grade Level

	Grade					
Performance on pretest	7	8	9	10	11	12
Below mean	0	0	52	41	9	1
Above mean	15	29	33	13	1	0

In fact, students in lower grades (seventh and eighth) all score above the mean on the pretest, whereas students in higher grades (9th–12th) are more likely to score below the mean. This suggests that grade level is an important (and natural) predictor of how students are likely to perform on the pretest.

Based on this analysis, we divided the algebra groups into two subgroups based on their grade level (middle school algebra or high school algebra). This was equivalent to dividing by district since middle school students were in one district and high school students were in the other. Once again, we checked for normality in the reduced data set and statistical similarity between the reduced and the complete data sets in both the middle and high school algebra data sets. As seen in Table 6, the distributions of pretest are statistically normal for the middle school and for the high school algebra groups. Posttest scores are also statistically normal for the middle school posttest. Unfortunately, posttest scores for students taking algebra in high school do not appear to be normally distributed. As a result, statistical procedures that require data be normally distributed cannot be used to analyze results involving posttest scores for the high school algebra group.

Table 6

Normality and Mean Equivalency Tests, by Algebra Class, for the Complete and the Reduced Data Sets on Both the Pretest and the Posttest

	Pretest K-S statistic						Posttest K-S statistic						Mann-Whitney test							
	Complete sample			Reduced sample			Complete sample			Reduced sample			Pretest				Posttest			
Class type	*Z*	*n*	*p*	*Z*	*n*	*p*	*Z*	*n*	*p*	*Z*	*n*	*p*	*U*	n_1	n_2	*p*	*U*	n_1	n_2	*p*
Middle school algebra	1.331	45	.058	1.266	44	.081	0.690	45	.727	0.744	44	.638	986	45	44	.974	981.5	45	44	.944
High school algebra	1.210	161	.107	1.184	158	.121	1.531	160	.018[a]	1.511	158	.021[a]	12555	160	158	.917	12490.0	159	158	.931

[a]The K-S statistic is significant indicating that the sample distribution is significantly different from the normal distribution.

Finally, we analyzed the reduced sample data set of both the middle school and the high school algebra groups for statistical similarity to the complete data set on the demographic variables of interest. As shown in Table 7 (middle school algebra) and Table 8 (high school algebra), a chi-square analysis suggests that the complete and reduced samples are statistically identical for both groups.

Table 7

Comparison of Key Demographic Variables in the Complete and Reduced Data Sets for Middle School Algebra Students

Variable of interest	Value of chi-square statistic	Degrees of freedom	2-sided significance
Gender	0.010	1	.921
Weekly amount of video game play	0.079	4	.999
Ethnicity	0.026	5	1.000
Previous year's math grade	0.000	3	1.000
Difficulty of pretest	N/A[a]		
Effort made to do well on pretest	0.023	1	.879
Difficulty of posttest	0.001	2	1.000
Effort made to do well on the posttest	0.032	1	.857
Student's perception that pretest was important	0.041	3	.998
Student's perception that posttest was important	0.055	3	.997

[a]All students indicated the pretest was "easier than other tests."

Table 8

Comparison of Key Demographic Variables in the Complete and Reduced Data Sets for High School Algebra Students

Variable of interest	Value of chi-square statistic	Degrees of freedom	2-sided significance
Gender	0.011	1	.917
Weekly amount of video game play	0.025	4	1.000
Ethnicity	0.068	6	1.000
Previous year's math grade	0.018	4	1.000
Difficulty of pretest	0.192	3	.975
Effort made to do well on pretest	0.053	3	.997
Difficulty of posttest	0.074	3	.995
Effort made to do well on the posttest	0.003	3	1.000
Student's perception that pretest was important	0.048	4	1.000
Student's perception that posttest was important	0.004	4	1.000

Each of the other classes involved in the study was also analyzed for similarity on these same parameters of interest. Once again, a chi-square analysis suggests that the complete and reduced samples were statistically identical.

Based on this analysis, we used the reduced sample formed using our second culling procedure. We have included the high school students in the Algebra 1 class when an analysis does not involve posttest results. Due to the deviation from normality of this population's posttest results, however, we have excluded this subsample when analyzing the quality of the posttest, learning gains between pretest and posttest, and to estimate the effect sizes associated with a single 40-minute exposure to various instantiations of *Save Patch.* Posttest results for the remaining six different class types are used for all analyses in this study. From our analysis, we identified how the game might best be used in our efficacy study and in future classroom interventions. The next section of this paper presents those results.

Results

Given the exploratory nature of our investigation and the fact that we will use these results to inform a future efficacy study, we have set the level of significance at $\alpha = .1$ for the results we reported in this study.

Pretest and Posttest Technical Quality

As was the case on previous occasions (Vendlinski et al., 2010), the pretest and posttest used in this study demonstrated high levels of technical quality. The inter-item correlation was high for the pretest ($\alpha = .948$, $n = 349$) as well as for the posttest ($\alpha = .959$, $n = 191$). Percent correct scores on the pretest were also significantly correlated with percent correct scores on the posttest for the control group ($r = .974$, $n = 22$). These measures suggest the test is highly reliable. Finally, the significant correlation ($r = .471$, $n = 42$, $p = .002$) between the pretest scores of those students who received only instruction on the game mechanic (the baseline condition) and the level those students ultimately reached in the game as well as with self-reported math grades the previous year ($r = -0.390$, $n = 303$, $p < .001$) and with self-reported math grades on the previous report card ($r = -0.457$, $n = 300$, $p < .001$) of the entire sample suggest that the pretest is a good measure of math knowledge in general and the knowledge it takes to be successful in the game. The negative correlations with grade are expected since "A" = 1, "B" = 2, etc. In this case, almost a quarter of the variability in the game level a student ultimately reached was explained by their performance on the pretest.

One possible criticism of using an identical pretest and posttest is that students will learn from the pretest and that such learning would be incorrectly attributed to the treatment. The study design allowed us to measure such gains since the control condition played a math video game that was unrelated to rational number addition for the same amount of time as students in the treatment groups. Arguably, then, any pretest to posttest gains in the control group would be the result of learning from the pretest items. In fact, the percent correct actually *fell* from pretest ($M = .5297$, $SD = .2945$) to posttest ($M = .5213$, $SD = .2977$) for the control group. This change, however, is not statistically different $t(21) = -0.581$, $p = .567$) and allows us to conclude that gains from pretest to posttest are unlikely attributable to merely taking the pretest.

Pretest Scores by Class Type

Before game play, each student took the pretest. Descriptive statistics for the pretest, by class type, are given for the reduced sample in Table 9 below.

Table 9

Mean Score on the Pretest by the Type of Class a Student Was Taking

Class type	*n*	*M*	*SD*
Algebra 1 middle school	44	.8649	.08109
Algebra 1 high school	158	.4787	.21953
Pre-algebra	45	.3564	.1887
Sixth grade math	25	.5623	.2133
Intro to Algebra	18	.5551	.2230
Algebra Success/CAHSEE[a]	23	.4151	.2133
Algebra Success/Algebra[a]	36	.4470	.2133

[a]Students taking either of the Algebra Success classes were counted only as part of their respective Algebra Success class and were not counted as part of the Algebra 1 high school class.

After taking the pretest, students were randomly assigned to play either one of the treatment versions of *Save Patch* or the control video game. While there were slight variations in the amount of game play in each class due to school schedules, students generally played for approximately 40 to 45 minutes in each class. Students were then asked to take a posttest.

Learning Gains Associated With Playing Any Version of *Save Patch*

To determine the learning gains associated with approximately 40 minutes of playing *Save Patch*, we compared the pretest and posttest means for the students in any of the treatment conditions. Student scores increased by approximately 1 percentage point from the pretest (M = .5451, SD =.2605) to the posttest (M = .5533, SD = .2690), but a paired samples t test suggests that these gains were not significant, t(168) = 1.458, p = .147.

Learning Gains Associated With Playing Particular Versions of *Save Patch*

Given these gains from pretest to posttest, we investigated whether the type of instruction within any of the various treatments was associated with significant pretest to posttest learning gains. Each of the interventions, the number of students assigned to that intervention (degrees of freedom), and the significance of pretest to posttest changes are given in Table 10 below.

Table 10

Pretest to Posttest Differences Between Individuals in Various Instruction and Feedback Conditions

Instruction/Feedback condition	Pretest		Posttest					
	M	*SD*	*M*	*SD*	*df*	*t*	*p*	d_z[a]
Graphics-based game mechanic instruction (baseline)	.5922	.2744	.6199	.2715	27	3.491	.002***	0.65
Graphics-based game mechanic instruction with video-based feedback	.5320	.2694	.5259	.2727	40	-0.490	.627	-0.08
Graphics-based math instruction with graphics-based feedback	.5219	.2566	.5243	.2693	38	0.176	.861	0.03
Video-based math instruction with graphics-based feedback	.5622	.2676	.5816	.2782	31	1.736	.093*	0.31
Video-based math instruction with video-based feedback	.5305	.2417	.5352	.2551	28	0.345	.733	0.06
Control (game played did not involve rational numbers)	.5297	.2945	.5213	.2977	21	-0.581	.567	-0.12

[a]Effect size is corrected for correlation between measures (G* Power).
*$p \le .1$. **$p \le .05$. ***$p \le .01$.

While the pretest to posttest gains associated with playing the *Save Patch* game, in general, did not appear to be significant, even at the α = .1 level, the results in Table 10 suggest that two of the instructional interventions are significantly associated with strong pretest to posttest learning gains at or below this level. In fact, the strongest results suggest that limiting instruction to how to play the game (i.e., just the game mechanics) produced a very significant pretest to posttest change that was either not evident, or was only marginally significant in the interventions that involved overt math instruction and feedback.

Learning Gains Associated With Playing *Save Patch* in Different Classes

We also analyzed the pretest to posttest differences associated with playing any version of *Save Patch* based on the math class each student was enrolled in. Surprisingly, given their high pretest scores, only the middle school algebra students who played the game made significant pretest (*M* = .8600, *SD* = .0825) to posttest (*M* = .8790, *SD* = .0822) gains, t(38) = 2.512, *p* = .016. These gains represent an effect size (corrected for pretest–posttest correlation) of 0.40 for this population.

Class Type and Instructional Interventions Interactions

In order to investigate the interactions between the type of class in which a student was enrolled and the various instructional methods used in the game, we conducted paired sample student *t* tests on each of these groups. As shown in Table 11, the gains made from pretest to posttest were significant for three of the groups and all the effect sizes were very large.

Table 11

Significance of Pretest to Posttest Gains (Paired Samples) for Students in Various Treatment Conditions by Class Type

	Pretest		Posttest					
Instruction/feedback condition and class type	*M*	*SD*	*M*	*SD*	*df*	*t*	*p*	d_z[a]
Graphics-based game mechanic instruction (baseline) in middle school algebra	.8527	.0985	.8807	.0901	8	3.404	.009***	1.14
Graphics-based game mechanic instruction (baseline) in sixth grade math	.5507	.2411	.5852	.2309	3	5.672	.011**	3.39
Video-based math instruction with graphics-based feedback in high school pre-algebra	.5147	.2563	.5803	.2592	7	4.146	.004***	1.47

[a]Effect size is corrected for correlation between measures (G* Power).
*$p \leq .1$. **$p \leq .05$. ***$p \leq .01$.

Pretest to Posttest Learning Differences

A key goal in this study was to prepare for a future efficacy study by: (a) determining which intervention(s) might produce the greatest differences in student learning; and (b) approximating the effect size of each identified intervention (see Research Question 2). Since different interventions seemed to be more or less effective depending on the math class a student was taking and, aside from sixth grade, that math class seemed strongly correlated with pretest score, we analyzed the effects of each different intervention, by class type, after controlling for pretest score.

As might be suspected from the fact that students only played the game for approximately 40 minutes between taking the pretest and the posttest, we expected the two tests to be highly correlated for each of the groups. This was indeed the case; however there was no significant pretest by treatment group interaction, $F(5, 179) = .138$, $p = .983$, and pretest scores were statistically similar across grades. Consequently, we used an Analysis of Covariance (ANCOVA) to control for pretest and to estimate the significance of differences between interventions within a particular class type.

The ANCOVA analysis suggests that none of the differences between treatment groups, after controlling for the pretest, were significant, F(5, 184) = 1.190, p = .316. Only one class exhibited significant between-group treatment effects, F (5, 38) = 2.305, p = .063. As shown in Table 12 and Table 13 below, an ANCOVA analysis does suggest that there were significant effects by instructional intervention for the high school pre-algebra students and that the most effective intervention for these students is video-based math instruction with graphics-based feedback.

Table 12

Pretest and Posttest Mean Scores and Standard Deviations for High School Pre-Algebra Students as a Function of Instructional Intervention

	Pretest		Posttest	
Instructional intervention	*M*	*SD*	*M*	*SD*
Graphics-based game mechanic instruction (baseline)	.2561	.0572	.2835	.0460
Graphics-based game mechanic instruction with video-based feedback	.3719	.1835	.3651	.1971
Graphics-based math instruction with graphics-based feedback	.3408	.1770	.3463	.1697
Video-based math instruction with graphics-based feedback	.5147	.2563	.5803	.2592
Video-based math instruction with video-based feedback	.3269	.1673	.2971	.1215
Control (game played did not involve rational numbers)	.2391	.0298	.2063	.0753

Table 13

Analysis of Covariance of High School Pre-algebra Students' Posttest Knowledge as Function of Game Instructional Format With Pretest Knowledge as Covariate

Source	*df*	*SS*	*MS*	*F*	*p*	η^2
Pretest (covariate)	1	1.056	1.056	218.978	< .001	.852
Condition	5	0.056	0.011	2.305	.063	.233

In fact, an analysis of the data suggests that students in the video-based math instruction with graphics-based feedback group perform significantly better, t(7) = 2.883, p =

.006, than students in any other group. These results, however, must be considered in light of the fact that the mean of pre-algebra students in the video-based math instruction and graphics-based feedback differed significantly from their peers in the other treatment conditions on pretest scores, $F(5, 39) = 2.111$, $p = .085$. While students in the video-based math instruction with graphics-based feedback condition gained significantly more from pretest to posttest than did their peers, as suggested above, they also displayed significantly greater understanding of the topic before playing the game.

This might suggest that, in general, students who have better pretest scores benefit more from *Save Patch* than students with lower pretest scores. We did not, however, find evidence of such an interaction in our data. We also investigated whether some basic level of understanding, as evidenced by pretest score, is necessary in order to show learning gains after playing the *Save Patch* game. Here again, a further analysis of the data does not support the notion that students who score above the mean (or various other thresholds) on the pretest benefit more from the game than students who score below such a threshold score. In fact, there does not seem to be a minimum pretest score that predicts learning gains in any of the treatment groups.

Correlation Between Game Level Achieved and Posttest Score

We next considered the relationship between how far a student progressed in the game and that student's score on the posttest to see if there was a correlation. Given the large number of game levels (50+) and the fact that students, on average, complete about 20 levels, we treated the maximum level achieved as a scale measure and computed Pearson's product-moment correlation coefficient to gauge the correlation between these variables. As expected, the correlation between how far a student progressed in the game and the student's posttest score is strongly correlated ($r = .433$, $p < .001$), and these correlations seem consistent across game instruction/feedback treatments as seen in Table 14 below.

Table 14

Pearson's Product Moment Correlation Coefficient (*r*) Between the Maximum Level a Student Reached in *Save Patch* and Posttest Score by instruction / feedback condition

Condition	*n*	*r*
Graphics-based game mechanic instruction (baseline)	28	.390*
Graphics-based game mechanic instruction with video-based feedback	41	.438**
Graphics-based math instruction with graphics-based feedback	38	.437**
Video-based math instruction with graphics-based feedback	31	.579***
Video-based math instruction with video-based feedback	29	.351*

*$p < .05$. **$p < .01$. ***$p < .001$.

When controlling for pretest using linear regression analysis, however, the maximum level a student achieved in the game is no longer a significant predictor of how the student will do on the posttest as shown in Table 15.

Table 15

Linear Regression Models Predicting Posttest Score Based on Maximum Level Reached in *Save Patch* and With Both Maximum Level Reached and Pretest Score

	Percentage correct on posttest		
		Model 2	
Variable	Model 1 *B*	*B*	95% CI
Maximum level reached	.013***	.001	[.000, .002]
Pretest percent correct		.979***	[.932, 1.027]
R^2	.188***	.927***	
F	38.143***	1046.406***	
ΔR^2		.740	
ΔF		1669.06***	

*** $p < .001$.

Relationship Between Deaths Per Level and Intervention

As might be expected, there are significant correlations between how far a student was able to get in the game and the average number of failed attempts (deaths) the student made per level ($r = -.458$, $n = 309$, $p < .001$). Not surprising, given the lack of instruction, the intervention with the highest number of deaths per level, on average, is the graphics-based game mechanic (baseline) version ($M = 1.381$, $SD = 1.028$). What is surprising, however, is that this lack of instruction did not seem to result in significant differences in the average number of deaths per level, t(307) = 1.585, p = .114, between these students and their counterparts who played one of the other more instructionally rich versions of *Save Patch* (M = 1.168, SD = 0.7682). Even with minimal instruction we found no significant differences in the maximum level attained between the *Save Patch* intervention groups, F(4, 304) = 0.325, p = .861, after 40 minutes of game play.

Conclusions

> Research Question 1: Can a video game be designed that helps students learn important mathematical concepts using minimal classroom time?

This study suggests that designing a video game with the goal of teaching important mathematical concepts is possible, even if the concepts have proven to be difficult for students to master in the past. In this initial study, we analyzed a video game designed, from its inception, to teach students how to add rational numbers. The design focused on two key foundational concepts, namely: (1) that the size of a rational number is relative to how one whole unit is defined; and (2) that addition allows us to combine identical units (or identical pieces of units) into a single sum. Rather than being added on as an afterthought after the game was designed, these foundational concepts were designed into the game mechanic at the outset, so that the game itself focused on these very specific learning objectives.

Our findings suggest that students using a game designed in this manner can increase their ability to add rational numbers even when playing the game for a relatively short period of time. In this study, the students played for only about 40 minutes, which is a little less than one class period. Given that understanding rational numbers and how to apply mathematical operations to such numbers is a longstanding, national shortcoming in American education, this is an important finding that seems most applicable to students in middle school or who are preparing to take algebra in middle school or high school.

> Research Question 2: Do different treatments of video game instruction or feedback produce different effects on student learning?

First, in our control condition (i.e., a video game intentionally designed to meet other learning objectives), students did not show significant learning improvement on the desired math content.

Second we generally did not find any significant learning gains for the game treatments when the game included mathematics instruction or feedback. However, we found significant learning gains between pretest and posttest when the game provided instruction only on *how* *to play* the game, that is, the version of the game that provided no overt math instruction or feedback. Other treatments of the game that included graphics-based and video-based instruction or feedback were generally not associated with significant student learning gains. This raises an important question.

Intuitively, it would seem that because all treatments included basic instruction on how to play the game that each treatment would produce a significant, positive learning effect. They did not. We hypothesize that while students in the minimal instruction condition seemed, on average, to fail at levels more often before passing them, such failure may have actually helped their learning. Why?

Unlike students in the feedback versions of the game who were eventually directed to complete the level in a certain way after a certain number of failures, students in the no math instruction version of the game had to solve each level on their own or give up on the game. Because there were no significant differences between the various treatment groups in how far students made it in the game, we believe that not only did students in the non-instruction and non-feedback group learn on their own, but also that it did not take any longer for them to do so. This would seem to support Charsky and Ressler's (2011) admonition that educators, "not dilute the potential effectiveness of games by taking away the one distinct attribute that gives them their advantage: play" (p. 614). We will analyze the preceding hypothesis in our full efficacy study.

Research Question 3: Is a one class period interaction with the game adequate to produce average student outcomes on the posttest that are commonly viewed as acceptable (i.e., greater than 70% correct)?

Even though students exposed to the treatment in this study played for a relatively short amount of time, only about 40 minutes, the learning gains in the version of the game that provided no math instruction were significant, and the effect size of the intervention proved moderate ($d = .65$). We believe this change is impressive given that the addition of positive rational numbers is generally taught in the fourth grade in California and, based on pretest scores, many of the students in this study seemed to be struggling with concepts that they were expected to have mastered two to eight years prior to this study. Nevertheless, given that students in the intervention that reported the largest gains still only averaged 62% correct on the posttest, additional treatments may be required in order to see larger, more acceptable, effects. This belief is supported by the strong correlation between the game level a student achieved and the student's posttest score, which suggests that it might be very important that a student play until achieving a certain stage in the game rather than merely playing until a set amount of time has expired. Because our experience suggests that the maximum time dosage a student can tolerate is about 40 minutes, it may also be important to spread this play to criterion over the course of several days. We plan to conduct studies that test such multiple treatments between the pretest and the posttest prior to our efficacy study.

Research Question 4: Do different treatments of videogame instruction or feedback produce differential effects for different types of students?

In addition to considering the length of time students play the game, our initial study suggests that a certain group of students might benefit more from playing a version of the game with more instruction and feedback than was beneficial for middle-school students who were studying math at grade level. Middle school algebra and sixth grade students—students at grade level in math—seemed to benefit more if they played the game without instructional priming or feedback. On the other hand, high-school pre-algebra students—students approximately two years below grade level in math—seemed to benefit most from a combination of video instruction designed to help them incorporate math concepts into game play and then text-based feedback if they struggled to correctly apply those math concepts in the actual game.

We noted, however, that the pre-algebra students in this instructional group scored considerably higher on the pretest than their peers who played other versions of the game. Consequently, we suspected that there may be some minimal level of understanding of rational number addition required before game play to benefit from playing the game. While pretest score was correlated to maximum game level achieved, we could find no minimal pretest score that seemed to serve as such a threshold. We also noted that the Introduction to Algebra class had a pretest mean that was statistically the same as the high-school pre-algebra students (and the sixth graders), and yet did not seem to benefit from this or any other version of the game in the same way.

Research Question 5: What other research questions should be answered prior to the full efficacy study?

We believe that this initial study suggests other research that might help improve our larger efficacy study. For example, it would be helpful to compare our "how to play the game only instruction" treatment to more standard "business as usual" conditions, such as textbook-based homework assignments or worksheets (Lee, Luchini, Michael, Norris, & Soloway, 2004). By design, creating or adapting game levels in *Save Patch* is straightforward and requires little time. Consequently, giving a random group of students game levels that are identical in content to what other students are receiving in paper-based homework or classwork exercises seems a logical comparison.

A number of students have expressed their preference for “playing the game rather than doing homework.” As such, requiring students to “play” a preset number of levels for “homework” may be a beneficial function that games can serve. A second, and more important line of research concerns the students who become stuck in a game like *Save Patch*, and who are unable to resolve the impasse on their own. Given that a “no instruction, no feedback” version of the game produces significant learning gains, what happens to students who become frustrated or paralyzed in such a condition? How might they be helped or their learning scaffolded to overcome such hurdles and achieve like their peers? Here again, we plan to address this question in a study prior to our larger efficacy study.

References

Baddeley, A. D. (1999). *Essentials of human memory.* Hove, England: Psychology Press.

Beesley, A. D., Apthorp, H., Clark, T. F., Wang, X., Cicchinelli, L. F., & Williams, J. M. (2011). *Classroom Assessment for Student Learning: The impact on elementary school mathematics in the Central Region* (NCEE 2011-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

California State Board of Education. (1997). *Mathematics content standards for California public schools, kindergarten through grade twelve*. Retrieved from http://www.cde.ca.gov/be/st/ss/documents/mathstandards.pdf

California State Board of Education. (2010). *Preliminary report of actions*. Sacramento, CA: Author.

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (2000). *Cognitively guided instruction: A research-based teacher professional development program for elementary school mathematics*. Madison, WI: National Center for Improving Student Learning and Achievement in Mathematics and Science, University of Wisconsin–Madison.

Charsky, D., & Ressler, W. (2011). "Games are made for fun": Lessons on the effects of concept maps in the classroom use of computer games. *Computers & Education, 56*, 604-615.

Chung, G. K. W. K., et al. (2011). *The effects of instructional variations on math and game outcomes using a prototype game to teach fraction concepts*. Manuscript in preparation.

Delacruz, G. C. (2011). *Games as formative assessment environments: Examining the impact of explanations of scoring and incentives on math learning, game performance, and help seeking* (CRESST Report 796). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Flew, T., & Humphreys, S. (2005). *Games: Technology, industry, culture*. South Melbourne: Oxford University Press.

G*Power (Version 3.1.3) [Computer Software]. Kiel, Germany: Universität Kiel.

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., et al. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (No. NCEE 2011-4024). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy.* New York, NY: Palgrave Macmillan.

Hamlen, K. R. (2011). Children's choices and strategies in video games. *Computers in Human Behavior, 27*, 532-539.

Kaiser Family Foundation. (2002). *Key facts: Children and video games.* Available from http://www.kff.org/entmedia/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=14092

Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & Education, 55*(2), 427-443.

Lee, J., Luchini, K., Michael, B., Norris, C., & Soloway, E. (2004, April). *More than just fun and games: Assessing the value of educational video games in the classroom. Proceedings of the Conference on Human Factors in Computing Systems, CHI 2004,* 1375-1378.

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.

Misquitta, R. (2011). A review of the literature: Fraction instruction for struggling learners in mathematics. *Learning Disabilities Research & Practice, 26*(2), 109–119.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Phelan, J. C., Choi, K., & Vendlinski, T. P. (2011, April). *The effects of a formative assessment intervention on student understanding of basic mathematics principles*. Poster presented at the annual American Educational Research Association conference, New Orleans, LA.

Phelan, J. C., Choi., K. C., Vendlinski, T. P., Baker, E. L., & Herman, J. L. (in press). Differential improvement in student understanding of mathematical principle following formative assessment intervention. *Journal of Education Research*.

Siebert, D., & Gaskin, N. (2006). Creating, naming, and justifying fractions. *Teaching Children Mathematics, 12*(8), 394-400.

Sweller, J. (1999). *Instructional design in technical areas.* Camberwell, Australia: ACER Press.

U.S. Department of Education. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: Author.

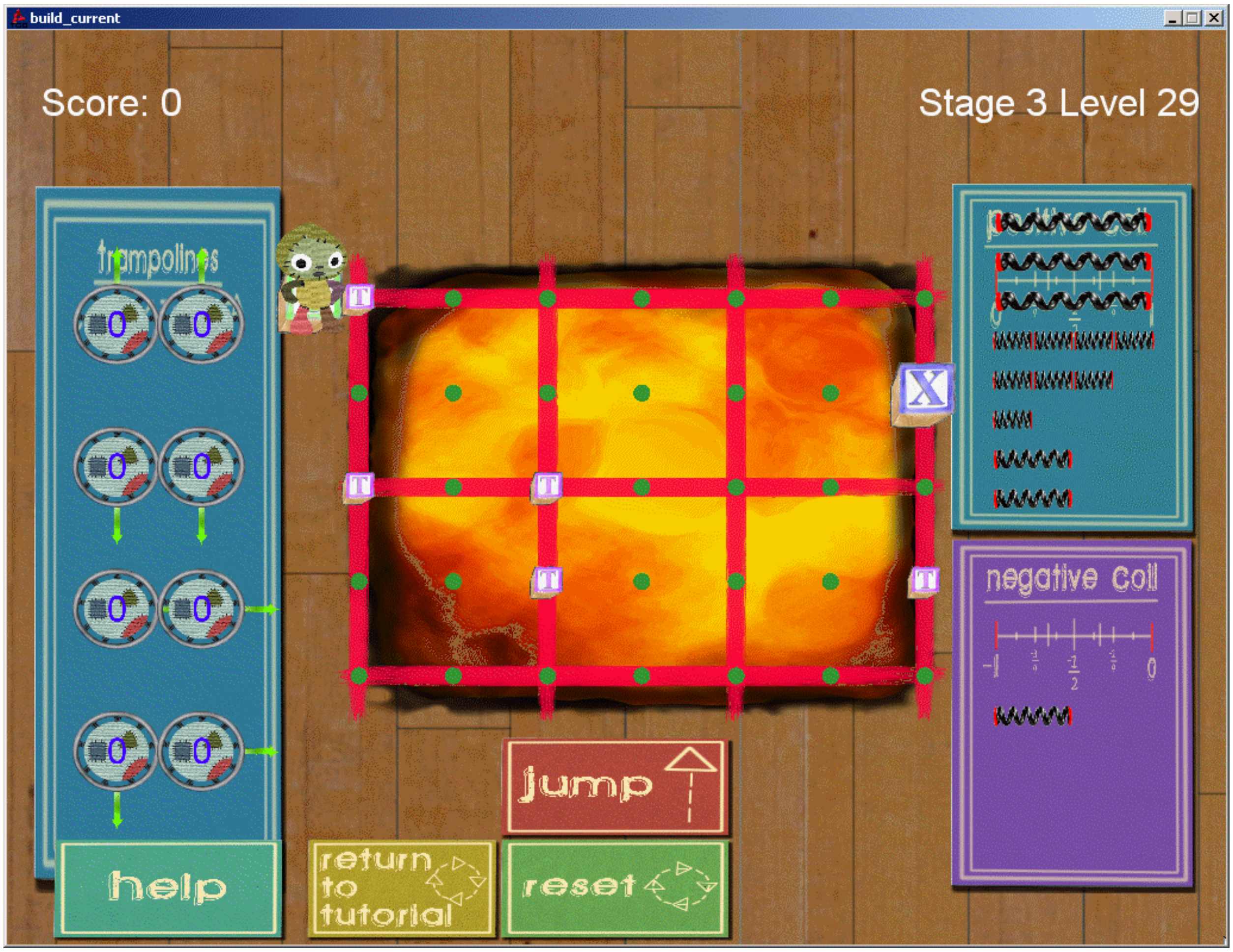
Vendlinski, T. P. (2011, April). Alternatives to repeating high school algebra. In P. Paek (Chair), *Putting policy into practice: How is eighth-grade algebra working?* Presentation to the 2011 Annual Meeting of the National Council of Teachers of Mathematics, Indianapolis, IN.

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 No. 774). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Vendlinski, T. P., & Phelan, J. (2011). *Using key conceptual ideas to improve teacher use of formative assessment data* (CRESST Report 794). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Appendix A:

***Save Patch* Game Board**



Appendix B:

Knowledge Specifications (Learning Objectives) in *Save Patch*

1.0.0 Does the student understand the importance of the unit whole or amount?

- 1.1.0. The size of a rational number is relative to how one Whole Unit is defined.
- 1.2.0. In mathematics, one unit is understood to be one of some quantity (intervals, areas, volumes, etc.).
- 1.3.0. In our number system, the unit can be represented as one whole interval on a number line.
  - 1.3.1. Positive integers are represented by successive whole intervals on the positive side of zero
  - 1.3.2. The interval between each integer is constant once it is established.
  - 1.3.3. Positive, non-integers are represented by fractional parts of the interval between whole numbers.
  - 1.3.4. All Rational Numbers can be represented as additions of integers or fractions.

2.0.0 Does the student understand the meaning of addition?

- 2.1.0. To add quantities, the units (or parts of units) must be identical.
  - 2.1.1. Identical (or common) units can be descriptive (e.g. apples, oranges, and fruit) or they can be quantitative (e.g. identical lengths, identical areas, etc.).
  - 2.1.2. Positive integers can be broken (decomposed) into parts that are each one unit in quantity. These single (identical) units can be added to create a single numerical sum.
  - 2.1.3. Each Whole Unit or part of a Whole Unit (fractions) can be further broken into smaller, identical parts, if necessary.
- 2.2.0. Identical (common) units can be added to create a single numerical sum.
- 2.3.0. Dissimilar quantities can be represented as an expression or using some other characterization, but are not typically expressed as a single sum [NB: we are considering numbers like 2 ¾ to have an implied addition – so 2 + ¾ – whereas 11/4 is a single sum].
- 2.4.0. Zero can be added to any quantity. When zero is added to any quantity, the value of the quantity remains unchanged (Additive Identity).
- 2.5.0. Adding two positive numbers will always produce a sum that is greater (more positive) than either number.
- 2.6.0. Adding two negative numbers will always produce a sum that is less than (more negative) either number.
- 2.7.0. Since they are opposites, adding a number and its opposite (two numbers of the same absolute value but opposite in sign) will result in a sum of zero (the additive inverse).

3.0.0 Does the student understand the meaning of the denominator in a fraction?

- 3.1.0. The denominator of a fraction represents the number of identical parts in One Whole Unit. That is, if we break the One Whole unit into "x" pieces, each piece will be "1/x" of the One Whole Unit.
- 3.2.0. As the denominator gets larger, the size of each fractional part (relative to the whole) gets smaller.

- 3.3.0. As the size of each fractional part gets smaller, the number of pieces in the whole gets larger.

4.0.0 Does the student understand the meaning of the numerator in a fraction?

- 4.1.0. The numerator of a fraction represents the number of identical parts that have been combined? For example, ¾ means three pieces that are each ¼ of One Whole Unit.
  - 4.1.1. If the numerator is smaller than the denominator, the fraction represents a number less than one whole unit.
- 4.2.0. If the numerator is equal to the denominator, the fraction represents one whole unit.
- 4.3.0. If the numerator is greater than the denominator, the fraction represents more than one whole unit.

5.0.0 Does the student understand any rational number can be written using fractions?

- 5.1.0. The numerator is the top number in a fraction
- 5.2.0. The denominator is the bottom number in a fraction.
- 5.3.0. Any rational number can be written as a fraction that relates one integer—the number of parts there are (numerator)—to another integer—the number of parts in one whole (denominator).
- 5.4.0. Proper fractions have numerators less than the denominator.
- 5.5.0. Improper fractions have numerators greater than or equal to the denominator.
- 5.6.0. Fractions where the numerator and denominator are equal represent One Whole Unit.

Appendix C:
Pretest to Posttest Effect Sizes (Uncorrected for Correlation) by Intervention and Class Type

Class type	Instruction condition	*M*	*N*	*SD*	Mean difference	*df*	*SD* pooled	Cohen's *d*
Pre-algebra	Gamey instruction, video FB	0.3719	11	0.18347	-0.01	20	0.19	-0.04
Pre-algebra	Math text instruction and FB	0.3408	10	0.17701	0.01	18	0.17	0.03
Pre-algebra	Video instruction and text FB	0.5147	8	0.25634	0.07	14	0.26	0.25
Pre-algebra	Video instruction and FB	0.3269	6	0.16732	-0.03	10	0.15	-0.2
Pre-algebra	Baseline (gamey)	0.2561	4	0.0572	0.03	6	0.05	0.53
Pre-algebra	Control (Mathemagic)	0.2391	6	0.02981	-0.03	10	0.06	-0.57
6th grade math	Gamey instruction, video FB	0.4593	4	0.14551	-0.07	6	0.15	-0.48
6th grade math	Math text instruction and FB	0.6057	5	0.25199	0	8	0.27	0.01
6th grade math	Video instruction and text FB	0.561	5	0.22651	0	8	0.22	0.01
6th grade math	Video instruction and FB	0.5366	4	0.14187	0	6	0.18	-0.02
6th grade math	Baseline (gamey)	0.5507	4	0.24112	0.03	6	0.24	0.15
6th grade math	Control (Mathemagic)	0.6789	3	0.34522	0.05	4	0.3	0.18
Introduction to Algebra	Gamey instruction, video FB	0.6889	2	0.11779	-0.07	2	0.1	-0.68
Introduction to Algebra	Math text instruction and FB	0.5345	4	0.23154	0.03	6	0.26	0.13
Introduction to Algebra	Video instruction and text FB	0.5122	3	0.27567	-0.01	4	0.3	-0.04
Introduction to Algebra	Video instruction and FB	0.3753	3	0.04066	0.03	4	0.08	0.38
Introduction to Algebra	Baseline (gamey)	0.8035	3	0.15806	0.03	4	0.17	0.16
Introduction to Algebra	Control (Mathemagic)	0.4675	3	0.24217	0.01	4	0.21	0.04
Algebra Success/CAHSEE	Gamey instruction, video FB	0.4476	5	0.19653	0.04	8	0.18	0.22
Algebra Success/CAHSEE	Math text instruction and FB	0.4857	4	0.34301	-0.03	6	0.32	-0.1

Class type	Instruction condition	*M*	*N*	*SD*	Mean difference	*df*	*SD* pooled	Cohen's *d*
Algebra Success/CAHSEE	video instruction and text FB	0.4532	4	0.26574	0.01	6	0.29	0.04
Algebra Success/CAHSEE	video instruction and FB	0.3659	4	0.15458	0.02	6	0.17	0.13
Algebra Success/CAHSEE	baseline (gamey)	0.2683	3	0.05589	0.07	4	0.09	0.75
Algebra Success/CAHSEE	control (Mathemagic)	0.4281	3	0.22147	-0.01	4	0.22	-0.05
Algebra Success/Algebra	gamey instruction, video FB	0.4612	11	0.28232	0	20	0.28	-0.02
Algebra Success/Algebra	math text instruction and FB	0.4236	9	0.15156	-0.02	16	0.16	-0.11
Algebra Success/Algebra	video instruction and text FB	0.3561	5	0.25947	0.01	8	0.28	0.05
Algebra Success/Algebra	video instruction and FB	0.4954	4	0.17598	-0.01	6	0.16	-0.08
Algebra Success/Algebra	baseline (gamey)	0.4927	5	0.19178	0	8	0.19	-0.01
Algebra Success/Algebra	control (Mathemagic)	0.4898	2	0.17522	-0.06	2	0.16	-0.38
Middle school algebra	gamey instruction, video FB	0.8994	8	0.04689	0.01	14	0.05	0.28
Middle school algebra	math text instruction and FB	0.8606	7	0.03918	0.02	12	0.04	0.55
Middle school algebra	video instruction and text FB	0.8484	7	0.10277	0	12	0.11	0.01
Middle school algebra	video instruction and FB	0.8384	8	0.10358	0.03	14	0.1	0.25
Middle school algebra	baseline (gamey)	0.8527	9	0.09853	0.03	16	0.09	0.3
Middle school algebra	control (Mathemagic)	0.9032	5	0.06363	0	8	0.08	-0.06
High school algebra	gamey instruction, video FB	0.5134	49	0.23226	0	96	0.23	0
High school algebra	math text instruction and FB	0.3754	30	0.18973	0.01	58	0.2	0.03
High school algebra	video instruction and text FB	0.5084	25	0.21595	0	48	0.23	0.01
High school algebra	video instruction and FB	0.5277	25	0.21777	0.03	48	0.23	0.14
High school algebra	baseline (gamey)	0.4303	14	0.23878	0.01	26	0.24	0.04
High school algebra	control (Mathemagic)	0.4853	15	0.17715	0.01	28	0.19	0.03