CRESST REPORT 796

Girlie C. Delacruz

GAMES AS FORMATIVE ASSESSMENT ENVIRONMENTS: EXAMINING THE IMPACT OF EXPLANATIONS OF SCORING AND INCENTIVES ON MATH LEARNING, GAME PERFORMANCE, AND HELP SEEKING

JUNE, 2011

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

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

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

Girlie C. Delacruz
CRESST/University of California, Los Angeles

June, 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.

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

To cite from this report, please use the following as your APA reference: 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).

TABLE OF CONTENTS

GAMES AS FORMATIVE ASSESSMENT ENVIRONMENTS: EXAMINING THE IMPACT OF EXPLANATIONS OF SCORING AND INCENTIVES ON MATH LEARNING, GAME PERFORMANCE, AND HELP SEEKING

Girlie C. Delacruz
CRESST/University of California, Los Angeles

Abstract

Due to their motivational nature, there has been growing interest in the potential of games to help teach academic content and skills. This report examines how different levels of detail about a game's scoring rules affect math learning and performance. Data were collected from 164 students in the fourth to sixth grades at five after-school programs. The treatment conditions were randomly assigned within each setting and included a control group (played a different math game); three variations of scoring explanations (elaborated, minimal, and no scoring information); and combined elaborated scoring explanation with incentives to access additional feedback. The scoring explanation alone did not lead to better math learning. However, compared to the minimal-to-no scoring information variations, the combined treatment of the elaborated scoring explanation and incentive resulted in higher normalized change scores and, after controlling for pretest scores, higher posttest scores. Implications of the results identify attributes for learning games in mathematics.

Introduction

Do games teach academic content and skills? Why? Developers and scholars argue that games capture the player's attention and engage them in complex thinking and problem solving (Barab & Dede, 2007; Gee, 2003; Jenkins, 2009; Malone, 1980; Rieber, 1996; Shaffer, Squire, Halverson, & Gee, 2005). What does the evidence say? Earlier studies of the effectiveness of games have produced inconsistent results and have shed little light on how game design influences learning outcomes (Hays, 2005; Ke, 2009; Kirriemuir & McFarlane, 2004; Mor, Winters, Cerulli, & Björk, 2006). Moreover, the influence of learner characteristics, such as prior knowledge (de Jong & van Joolingen, 1998; Moos & Azevedo, 2008; Plaas et al., 2009; Shute, 1993) and self-efficacy (Backlund, Engström, Johannesson, Lebram, & Sjödén, 2008), has been shown to mediate learning in games. The findings investigating competitive versus cooperative learning environments have also indicated that playing cooperatively leads to better learning from games, regardless of individual differences (Ke & Grabowski, 2007).

The aim of this report is to investigate the impact of game design, especially its use of formative assessment. Games include measures of student progress and various types of immediate feedback or formative assessment. Analyzing student errors and giving high-quality feedback are hallmarks of formative assessment (Cizek, 2010; Popham, 2006; Taras, 2010; Wiliam, 2010). Like the research on games, however, there are very few empirical studies that have directly tested variations of these elements, with most of the empirical effort focused on teachers and classroom use.

The first purpose of this study is to examine how different levels of detail about a game's scoring rules affect math learning and performance. The second goal of this study is to examine the effect of providing an incentive to players to access additional feedback. The outcomes studied include math achievement, game performance, and help seeking by accessing feedback when needed. By using games as a context for formative assessment, findings from this study intend to strengthen understanding of feedback in effective game design.

Literature Review

This review will consider prior evidence in game design, formative assessment, and feedback and use evidence to develop the framework of the study. The first part of the review focuses on the research on games for learning. Traditional elements of games of leisure are contrasted with learning games. A range of rationales arguing for educational games is followed by a review of relevant empirical studies.

The second part of the review examines the practice of formative assessment—specifically, the features that scholars have cited as important for effective learning. The empirical evidence of these claims is also evaluated, focusing particularly on types and uses of feedback.

Games for Learning

Rationales for learning with games. One fundamental premise underlying the use of games for educational purposes is that they can be more engaging than traditional classroom activities (Malone, 1981; Rieber, 1996). Specifically, games appear to be intrinsically motivating, with the source of motivation and enjoyment originating from game play alone, rather than something external to the task (Malone, 1980). Academic engagement has been shown to impact learning (Fisher et al., 1978) and if games have the power to extend engaged learning, then they should have impact on achievement measures.

Driven by the idea that learning should be "fun," many of the earlier games that were developed for educational purposes, termed *edutainment*, simply overlaid conventional gaming principles with academic content (Buckingham & Scanlon, 2003; Egenfeldt-Nielsen, 2005; Okan, 2003). The intent was to persuade students to complete academic tasks, in two ways: (a) to work on academic tasks and be rewarded by playing games, or (b) to play a game in which the academic content to be learned was embedded in the game, though it was not intrinsic to the game play itself (Kebritchi, Hirumi, & Bai, 2010).

One example of this form of edutainment was *Math Blaster Episode 1: In Search of Spot*. The objective of this game was to collect "trash" in the universe, which was achieved by correctly solving a math problem to release a tractor beam that would collect the trash. The use of math was not intrinsic to the problem (such as determining the force of the beam or the distance the beam needs to travel). Instead the "doing of math" was extrinsic to the game. For example, correctly computing an addition problem was rewarded by game mechanics (release of tractor beams), which are "part of a game's rule system that covers one general or specific aspect of the game" (Boardgamegeek, 2005, as cited in Björk & Holopainen, 2005, pp. 413).

Drill-and-practice elements in games remain even in modern, highly immersive learning games. One example is *Dimension-M*, which is a video game that was designed to teach algebra. Successful completion of math tasks tangential to the game play itself (e.g., find all of the prime numbers in this space) is rewarded with the ability to move forward in the game. Nevertheless, games that employ drill and practice can be successful for the practice of existing skills. For example, research on the effectiveness of researcher-developed pre-algebra and algebra games in classrooms over an 18-week period demonstrated that playing the game leads to higher achievement on math measures (e.g., district-wide benchmark tests) than that of a control group (Kebritchi et al., 2010).

Some scholars have argued that the value of using games for learning extends beyond creating a "fun" practice environment because games encourage active participation by learners in relevant discourse practices thought to engender deeper, conceptual learning (Gee, 2003; Shaffer & Gee, 2007; Squire, 2006). For example, Shaffer and Gee (2007) assert that games are beneficial for learning because they can embody an *epistemic frame*, which they defined as a "community's distinctive ways of doing, valuing, and knowing" (pp.76). The idea of an epistemic frame is not unlike Gee's (2003) contention that games reflect a semiotic domain or "any set of practices that recruits one or more modalities (e.g., oral or written language, images, equations, symbols, sounds, gestures, graphs, artifacts, etc.) to communicate distinctive types of meanings" (pp. 18).

Currently, research evidence supporting claims that games affect complex learning is very limited. Some studies do suggest that games may facilitate appropriate discourse practices as students adopt professional roles and learn to use relevant conceptual tools and resources (Clark, Nelson, Sengupta, & D'Angelo, 2009; Rosenbaum, Klopfer, & Perry, 2007; Shaffer, 2006; Squire & Jan, 2007; Svarovsky, 2010). For example, findings from studies of the game *Environmental Detectives* suggest that when students work under similar conditions as professional environmental scientists, the constraints imposed guided systematic reflection about how best to approach the task (Klopfer & Squire, 2008). Likewise, Svarovsky (2010) used *Digital Zoo* in her work with middle school girls to help them develop engineering frames, which she identified as engineering skills (e.g., comparing alternatives), knowledge (e.g., use of professional terms), identity (e.g., engineer as innovator), values (e.g., importance of reliable design), and epistemology (ruling out a costly design). Playing the game led to an increased understanding of the knowledge of center mass as measured by analyses of interviews and conceptual physics textbook problems.

Yet, several studies do not report enough data to draw significant conclusions. For instance, after playing *Outbreak @ the Institute*, a game designed to teach students about the epidemiology of disease, it was reported that one of the students "adopted the habits of mind of a medical doctor in such a situation who might feel the obligation to cure all of the patients" (Rosenbaum et al., 2007, pp. 40). However, it is difficult to determine whether the game had the same impact on all of the students, as those data were not reported. Similarly, Squire and Jan (2007) focus on a case study of the three highest performing students in the study. While their rationale was that the particular group chosen provides a model for effective communication in the game, as the authors acknowledge, it is not clear whether the game was effective for all students, especially weaker readers.

The issue of game effectiveness is further complicated by the use of assessments such as standardized tests. Such tests may be convincing to school personnel but are often not appropriate for games that emphasize deeper learning of a few concepts (Clark, Martinez-Garza, Nelson, D'Angelo, & Slack, 2009; Hickey, Ingram-Goble, & Jameson, 2009). While research on the development of advanced statistical and measurement methods is ongoing, existing assessment methodologies are not sophisticated enough to fully capture the richness of the learning that may be occurring in these games (Iseli, Koenig, Lee, & Wainess, 2010; McCreery, Krach, & Schrader, 2010; Rupp, Gushta, Mislevy, & Shaffer, 2010; Shaffer et al., 2009) or their validity.

Evidence on the Effectiveness of Games for Learning

In general, meta-analyses or reviews of past research on the effectiveness of games for learning report inconsistent findings—even for games with similar topics (e.g., Cameron & Dwyer, 2005; Lee, Luchini, Michael, Norris, & Soloway, 2004; Papstergiou, 2009; Spivey, 1985). For instance, Ke (2009) reports that out of 65 studies reviewed, a little over half of the studies reported positive effects of the game; about a quarter of the games reported mixed results (with games being more effective for either some outcomes or certain learners); one study favored conventional instruction as more effective; and the rest reported either no difference between computer games and conventional instruction. Ke's findings were consistent with the outcomes of other reviews of games for learning (see de Freitas, 2006; Dempsey, Rasmussen, & Lucassen, 1994; Greenblat, 1973; Hays, 2005; Kirriemuir& McFarlane, 2004; Tobias & Fletcher, 2007; Vogel et al., 2006; Wainess, O'Neil, & Baker, 2005).

Recent studies of immersive and complex games have produced promising findings, especially in the area of science education (Clark et al., 2009). However, to date there is little evidence that in the domain of mathematics, games teach significant intended goals (Mor et al., 2006), such as those espoused by the emerging national consensus on the Common Core Standards in Mathematics and Literacy (Council of Chief State School Officers, 2010).

There may be several explanations for the inconsistency in findings. First, many of these studies compared one instance of a group, that is, one game of many possible games with one instance of "traditional practice." The methodological flaws in studies that compare technological media (e.g., computers) versus conventional instruction are well known (Clark, 1994), raising questions about the scientific generalization of their findings (Lumsdaine, 1963; Mielke, 1968).

Second, when there is a lack of variation in the treatments, for instance comparing a game with nothing, it is impossible to isolate the elements that are effective or ineffective. For example, Rosas et al. (2003) found that there was no difference in learning between students who engaged in Nintendo-like games designed for learning and those who received traditional instruction. However, the authors surmised that one explanation of the findings is that the control teachers were more careful in their teaching methods because they knew they were being evaluated. Similarly, Kebritchi et al. (2010) provided teachers with resources and training on how to incorporate pre-algebra games in the curriculum. The study's randomization process used intact classrooms, however, and their analyses of the data did not incorporate the variable of teacher implementation. Instead, the analyses focused on the student, therefore potential teacher implementation effects cannot be ruled out.

Additionally, Laffey, Espinosa, Moore, and Lodree (2003) conducted a study aimed at evaluating the potential of interactive computer technology to teach math skills to young, low-income, urban children. Both the treatment group and the control group received traditional classroom instruction. The students in the treatment group outperformed the control group in terms of differences between pretest and posttest scores. Still, it is difficult to draw conclusions from this study because students in the treatment group received extra time engaging in math tasks by playing commercially available math games, such as *Might Math* and *Millie's Math House*. Therefore, it is unclear whether or not higher achievement was attributed to playing the games or just engaging in more math tasks. In any case, the study suffers from the methodological flaw of conducting an evaluation of a pair of instances and attempting to generalize to the class; that is, games or classroom instruction.

Research on Game Design Features

In addition to inconsistent and methodologically flawed studies of entire games, there is little research on game features that lead to learning. For instance, Ke (2009) reported that only 17 out of the 89 studies that were reviewed focused specifically on elements of game design. Many of the design choices in these studies have been governed by cognitive load theory (Ayres & van Gog, 2009; Moreno, 2009; Paas, van Gog, & Sweller, 2010), multimedia learning principles (Clark & Mayer, 2007; Mayer, 2005), or features of social learning (Webb & Palincsar, 1996). For example, studies have examined how the specification of goals reduced cognitive load (Miller, Lehman, & Koedinger, 1999); how the presence of pedagogical agents influenced learning (Conati & Zhao, 2004; Moreno, 2004; Van Eck & Dempsey, 2002); and which group dynamic was more effective (Ke & Grabowski, 2007). While useful, aspects of game design that derive from instructional research remain largely understudied. For example, little is known about how the feedback and scoring rules influence game learning.

Formative Assessment

One defining characteristic of games is that they evaluate the player (Juul, 2000) and use interpolated assessment information as triggers for adapting the learner's experience. They accomplish this adaptation by giving explicit feedback or by using technology to modulate the challenges of the academic material and cognitive demands to sustain play and affect learning (Baker & Delacruz, 2007; Salen & Zimmerman, 2003; Schell, 2005).

This description of games is almost synonymous with the notion of formative assessment. Popham (2006) provides this definition:

> An assessment is formative to the extent that information from the assessment is used, during the instructional segment in which the assessment occurred, to adjust instruction with the intent of better meeting the needs of the students assessed. (pp. 3-4)

In general, the use of performance and process information (both by the student or the teacher) with the intent to improve learning underlies the idea of formative assessment (Airasian & Jones, 1993; Baker, 1974; Black & Wiliam, 1998a, 1998b, 1998c; Markle, 1967; National Research Council, 2001, 2003; Scriven, 1967). In a similar fashion, games are often equipped technologically to interpret game play (delays and corrections, for example) and to use such information to make critical instructional decisions. For example, the use of artificial intelligence in games makes it possible to maintain persistence when players are "thrashing," that is, not proceeding in an orderly way, by reducing the difficulty of the challenge (J. Hight, personal communication, July 15, 2009). Artificial intelligence can also be used to present information needed to succeed in a failed task, much like feedback or the instructional supports that are provided by a teacher.

Viewing games for learning as formative assessment environments introduces the importance of considering how to make the criteria of performance transparent for the player. It is often argued that explicitly defined criteria for evaluation are critical for effective formative assessment (Black & Wiliam, 1998a; 1998b; 1998c; Frederiksen & Collins, 1989; Sadler, 1998; Taras, 2010; Wiliam, 2010). This study is focused on learning games where game play is directly integrated and linked with its corresponding academic content. If game play is to require the use, demonstration, and evaluation of the application of knowledge or skills in a particular domain, game players may need to know the criteria underlying their progress or lack thereof.

Rubrics of Scoring Rules to Promote Clarity of Expectations

In formal learning settings, scoring rubrics (or codified scoring rules and corresponding score values) have been used to improve assessment clarity by making explicit what constitutes good performance. There have been numerous studies on the teacher use of rubrics, typically for student-constructed responses. Most of the research has examined how best to train teachers to effectively use rubrics to increase the reliability and validity of scoring performance assessments (Baker, Abedi, Linn, & Niemi, 1995; Gearhart, Herman, Novak, Wolf, & Abedi, 1994; Koretz, McCaffrey, Klein, Bell, & Stecher, 1993; Supovitz, MacGowan, & Slattery, 1997).

It has also been argued that in order for formative assessment to work, focus should be on how the student interprets and uses the given information (Popham, 2006; Stiggins, 2005). While student use of rubrics to improve learning appears logical, only a few studies have examined this idea directly. In Jonsson and Svingby's (2007) summary of research on the use of rubrics, only one third of the studies reviewed focused on student use. Moreover, those studies focused primarily on the impact of rubrics for peer review. It has been shown that rubrics can result in student judgments similar to that of classroom teachers (Brown, Glasswell, & Harland, 2004; Sadler & Good, 2006). Also, when compared to using rubrics to score the tests of others, students benefited more from the evaluation of their own tests (Sadler & Good, 2006).

However, interpretation of the results from the few empirical studies on the use of rubrics to guide performance of a task is inconclusive due to methodological flaws. Davidson and Scripp (1990) gave some participants a written rubric to facilitate the development of a classification system for a group of arthropods. While they found that compared to the control group, the treatment group had higher content learning and tended to refer to the criteria in the rubric during their self-evaluations of their work, the existence of a confounding variable makes it difficult to draw conclusions. Specifically, the treatment group was asked to stop in the middle of the classification task, engage in self-assessment, and then revise accordingly. In contrast, the control group was never asked to self-assess during their task.

Similarly, Mullen (as cited in Jonsson & Svingby, 2007) demonstrated that giving the rubrics as guides for lab write-ups led to an increase in the quality of the students' products. White and Frederiksen (1998) also found that providing rubrics to students was effective for learning concepts of force and motion. However, like Davidson and Scripp (1990), students in Mullen's (2003) and White and Frederiksen's (1998, 2000) studies engaged in an active process of self-assessment. Therefore, it is difficult to attribute increased achievement to the use of the rubrics alone. An alternative explanation may be that the act of engaging in self-assessment contributed to deeper processing and better learning.

The use of inadequate interventions also makes interpreting findings problematic. For example, Green and Bowser (2006) used rubrics to help communicate the criteria of a literature review to graduate students, but found no clear performance differences between students who received the rubric and students who did not. The authors surmise that two of the explanations for the results of their study were that the students were not skilled at using the rubric to guide their writing, and that the criteria on the rubric were not communicated as explicitly as was needed to improve performance. Brown et al. (2004) examined the effectiveness of a writing rubric to teach students with poor writing skills how to write persuasively. While the students in the study improved considerably, a within-subjects design was used. It is unclear whether the rubrics were effective because they supported self-assessment, clearly communicated evaluation criteria, or were given prior to beginning the task to guide performance.

Similarly, in non-game settings there is mixed evidence on the effectiveness of rubric or scoring clarity on student achievement. Methodological flaws include either confounding factors, use of inadequate interventions (unclear rubrics), or design errors. The practice of using rubrics to clarify understanding for markers is a worldwide phenomenon. It is possible that in the more controlled context of game play, explicit scoring rules will demonstrate their impact on student learning.

Feedback for Formative Assessment

A tenet of formative assessment requires effective use of feedback, so that students have a clear understanding of the learning goals or objectives (and the criteria that define good quality), a way to relate their own performance to that goal, and then a path to achieve the goal if there is any discrepancy between the goal state and their own performance (Black & Wiliam, 1998a; Heritage, Kim, Vendlinski, & Herman, 2009; Sadler, 1998; Taras, 2010; Wiliam, 2006). Feedback related to progress in learning tasks can generally have a positive effect on learning (Nyquist, 2003). However, there are certain conditions that make it effective (see for example, Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Hattie & Temperly, 2009; Kluger & DeNisi, 1996).

Examining the research on feedback, Shute (2008) suggested guidelines for conditions in which feedback can be used to support learning. Feedback, for instance, should be task-centered rather than based on normative comparisons, with the accuracy of performance judged against some established standard or quality. Feedback needs to help students understand the goals of the task, especially since evidence suggests that students and instructors often have different ideas about the goals of the task or the criteria for evaluation (Hounsell, 1997; Norton, 1990). Feedback should convey information rather than simply letting a student know whether he or she has been successful or not. Students also need elaborated information about the particular elements of the task.

Use of Feedback in Games

In games, when event data are used to make some judgment about performance and have consequential meaning for the player, feedback is usually given to the player. As such, one affordance of games is that they can be designed to provide instructional feedback that is responsive to specific actions and behaviors. The next section of the literature review explicates the conceptual underpinnings of the second prong of this study—motivating students to use provided instructional feedback help when faced with difficulties in the game, in other words, motivating students to engage in effective help seeking.

Germane to the discussion of student use of feedback is the research on help seeking. The type of help that is often used in help-seeking studies is tailored hints and feedback, or supporting information accessible in general help menus. To understand the relationship between help seeking and learning, researchers have primarily studied the role of cognitive factors (e.g., self-efficacy), the nature of the provided help (e.g., context-specific versus generalized principles), timing and frequency (e.g., on demand or system-initiated), and how the process of help seeking can be taught explicitly (for a review, see Aleven & Koedinger, 2003).

However, while it may seem rational that students would attempt to use or seek additional feedback or hints to overcome an impasse in performance, the research demonstrates that students either use ineffective help-seeking strategies, or avoid seeking help altogether. This is problematic as the lack of compliance with the treatment greatly reduces any potential effect on the outcomes. For example, Nelson (2007) compared different levels of feedback on student achievement in an immersive learning environment. Results from the study indicate that most students did not access the feedback. Also, between students who were provided extensive feedback and students who were given moderate feedback, there were no statistical differences on the frequency of accessing the hints. In a study that examined the effect of providing user-initiated feedback via a pedagogical agent, Van Eck and Dempsey (2002) also reported low levels of student access to the feedback. The authors concluded that further research is essential to figure out how to promote its use.

Results from these studies suggest that the mere presence of feedback, hints, or available help is insufficient for learning. One approach has been to train students in how to engage in effective help seeking (Salden, Aleven, & Renkl, 2007; Schwonke et al., 2007). Otherwise, in the literature on games, feedback, and help seeking, there is a noticeable gap in understanding how to motivate students to engage effectively in help seeking and to use provided feedback. The use of incentives for seeking help and feedback provides a relevant avenue for exploration.

Incentives for Learning

Use of incentives in formal learning settings. The evidence of the effectiveness of incentives such as reward on shaping behavior dates back to Skinner's (1956) notion of operant conditioning theory. As the number of studies on the relationship between incentives and motivation increased, the use of incentives for learning in formal settings became discouraged. Findings that incentives reduced intrinsic motivation, especially for tasks that had high initial interest, led scholars to conclude that external rewards were detrimental to learning (Deci, Koestner, & Ryan, 1999, 2001).

Yet, not all studies have produced findings that suggest a negative relationship between incentives and learning. Rather than using studies where the initial interest in the task was high, Cameron's (2001) meta-analysis focused on studies where initial interest in the task was low. Results indicated that there were no negative effects of rewards on intrinsic motivation when initial interest in the task was low (Cameron, 2001; Cameron & Pierce, 1994).

Although it has been argued that there is a direct relationship between intrinsic motivation and learning (Deci & Ryan, 2000), the conclusion that incentives are bad for learning is problematic. The studies used in Deci et al.'s (2001) and Cameron's (2001) meta-analyses focused on how rewards affected intrinsic motivation, not learning. As a result, these studies employed measures of intrinsic motivation (e.g., time on task, willingness to repeat task) as the dependent variables. While it is appropriate to study impact on learning processes, the lack of direct measures of learning outcomes makes the conclusion regarding the relationship between incentives and learning of less value.

Another related argument against the use of incentives for learning is that the incentives are only present temporarily, and that intrinsic motivation to do a task is greatly reduced once the incentive is taken away (Deci et al., 1999). In video games, incentives are intrinsically tied to performance and are often a permanent aspect of the activity. In fact, game designers argue that incentives (either in the form of rewards or punishment) are essential to the fun or sometimes the compulsion of the experience (Fullerton, 2008; Koster, 2005; Schell, 2005).

Using Incentives to Motivate Accessing Feedback

This discussion on the use of incentives for learning is relevant to the proposed study as a potential approach to motivating the access and use of feedback. That students avoid help seeking is not surprising given that historically, it has been discouraged in most learning contexts. In game-based settings, accessing feedback also slows the game down, and some games have speed of play as a basis of advancement.

In a similar vein, research conducted in classroom settings suggests that there is social stigma attached to help seeking, a public sign of failure (Karabenick & Newman, 2006; Puustinen, Vockaert-Legrier, Coquin, & Benicot, 2009; Ryan & Pintrich, 1997). Although

ILEs are generally more anonymous, accessing feedback is still associated with less rather than more proficiency. For example, when a hint is given to a student in the PACT Geometry tutor, the student's visible "skill bar" decreases (Aleven & Koedinger, 2000).

However, if one adopts the perspective that incentives can communicate what is valued within a given context and promote desired behavior, offering an incentive to use feedback may have a positive impact on learning. Rewards are, after all, a method for signaling those actions or behaviors that are encouraged or discouraged by a particular community. Conveying these expectations is a key process through which individuals learn how to participate in the communities that they are members of (Rogoff, 1990, 2003).

Summary

The review of this literature first defined a range of games types, and then focused on learning games. The evidence in support of learning game effectiveness is mixed; methodological limitations provide few answers about why games are effective in some contexts and not in others. There is a paucity of research on variations in instructional game design itself; thus, little is known about what makes games effective, for which learners, and under what conditions.

This study proposes an alternative perspective—approaching game design as if it were formative assessment design. The research on formative assessment in traditional learning settings, while still underdeveloped, suggests this approach may prove to be fruitful. Communication of a game's scoring rules can be leveraged to make assessment criteria more explicit and transparent. When the scoring rules are appropriately tied to academic progress, providing an explanation of the game's scoring rules functions as a rubric in a game for learning. The scoring rules make explicit the stated learning objectives of the game as well as the criteria used to evaluate performance. The scoring rules can direct attention to what and how responses are being scored which may (a) make more explicit the learning objectives or goals of the game, (b) clarify the criteria of performance (i.e., what "counts"), and (c) support the development of self-assessment of performance to determine when additional help is necessary. This information can be provided prior to game play to guide performance, or as a context for elaborated feedback.

The review of the research on the use of feedback in games demonstrates that students rarely access voluntarily available feedback. Furthermore, the focus of research studies on feedback has largely ignored how best to motivate students to use the feedback. The use of an incentive to access feedback may be one potential approach, especially as findings from studies using tasks with initial low interest suggest that incentives may be beneficial. At best, an incentive can reverse the association between seeking help and failure, signaling to the student that seeking feedback is a valuable act leading to proficiency.

Methods

This part of the report provides an overall description of the research questions and design of the study, and a description of the procedures, including (1) the game design and independent variables, (2) the dependent measures, (3) the selection and assignment of students, and (4) the data collection process.

Research Questions

1. What is the effect of a pre-algebra math game on learning outcomes of upper elementary students compared to a control group?
2. What is the effect of providing an explanation about the scoring rules on (a) math learning and (b) game performance?
3. Do incentives to seek additional feedback affect (a) math learning and (b) game performance?
4. Does a) lack of information or b) providing an incentive affect the frequency with which students voluntarily access feedback?

Hypotheses for the Math Learning and Game Performance Outcomes

1. Students who are given an explanation of the scoring rules will outperform students who are given minimal or no scoring information.
    a. Rationale: The explanation of scoring rules provides relevant and useful information regarding criteria and feedback of performance.
2. Students who are given both the explanation of the scoring rules and an incentive to access feedback will outperform (a) those given only the explanation of the scoring rules and (b) those who received minimal or no scoring information.
    a. Rationale: The incentive should motivate students to use provided feedback during the game to add to the benefit of having the elaborated scoring rules information.

Hypotheses for Accessing Feedback Help

1. Students who receive minimal scoring information will voluntarily access feedback more frequently than students who receive an explanation of the scoring rules.
    a. Rationale: The students who receive minimal scoring information will need to use the feedback most often because their lack of information should limit their understanding of the underlying causes of their mistakes.
2. Students who receive an incentive that returns lost points to the player will voluntarily access feedback more frequently than students who are given an explanation of the scoring rules but no incentive.

   a. Rationale: The incentive will motivate students to use the feedback help because points are an important aspect of the game.

3. Students who receive an incentive will voluntarily access feedback more frequently than students who receive minimal scoring information.
   a. Rationale: Compared to a lack of information, an incentive is a stronger motivation to access feedback.

Research Design

A randomized-control, 1 × 5 design (Kirk, 2009) was used in this study. There was a control group and four treatment conditions, which are described in the following section. Within participating after-school programs, each participant was randomly assigned to one of the five conditions.

Description of Conditions Used in the Study

Two pilot tests were conducted to refine the assessments and the final set of independent variables. Variations of explanations of scoring rules (e.g., the timing of the information—before game play or during the game) and the motivation of help seeking (e.g., competition, external prizes) were tested. As a result of the two pilot studies, the final set of independent variables used in the study were reduced to (1) variations in the amount of information provided to the player regarding the scoring rules and (2) incentives directly incorporated in the game to encourage effective help-seeking behavior. In this section, a description of all of the conditions used in the study will be provided. First, the control group will be described, followed by details of each individual treatment condition.

Control group. The control group played a different math game on an iPod Touch.[1] The goal of this math game was to find sets of matching and adjacent whole numbers. More points were awarded for larger sets. This game was chosen because it was in the domain of math but did not target the same learning objectives as the experimental game.

Treatment Conditions

Players in all of the treatment conditions played games that contained tutorials, feedback, a general help menu, and opportunities to access additional feedback.[2] The treatment conditions differed along two dimensions based on the independent variables: (1) amount of information provided regarding the scoring rules, and (2) the use of incentives to

---

[1] While it may be expected that students in the control group would be more motivated to play given the different modes of technology (i.e., iPod Touch versus a laptop), this was not the case. When provided the opportunity to continue game play, most students who played the iPod Touch game wanted to try the laptop game.

[2] Details of these components of the games will be elaborated in the section on game design.

motivate accessing instructional feedback help when a mistake was made. Table 1 contains summary descriptions of the conditions. Following the table, each of the conditions is described in more detail.

Table 1

Treatment Condition Descriptions

	Independent variables				
	Amount of information player was given regarding game's scoring rules				
Treatment condition	Explanation of scoring rules given prior to game play	Explanation of scoring available in general help menu	Explicit feedback about when and how many points are earned and lost	Math elaborated feedback of scored event	After a mistake was made, player was given an incentive to access additional feedback help
No scoring rules info.	O	O	O	O	O
Points-only feedback	O	O	X	O	O
Explanation of scoring rules	X	X	X	X	O
Rewarding help seeking	X	X	X	X	X

No scoring rules information. In this condition, no scoring rules information was given other than (a) information on where the points were displayed on the screen, and (b) change in the number of points displayed at the completion of the level. The information about points earned was presented at a summary level, so players in this condition did not benefit from the knowledge of which events triggered either the attainment or loss of points. The students in this condition played a version of the game that provided no feedback beyond knowledge of results. For example, when a student failed to complete a level, the character in the game exploded. No explicit information, however, was given regarding the potential cause for the mistake. It was expected that this would be the most difficult version of the game to play.

Points-only feedback. This treatment condition was similar to the no scoring rules information condition with one exception. Players in the points-only feedback condition were alerted explicitly when an event that triggered the earning or losing of points occurred.[3] For instance, when a successful move was made, players were presented with a feedback screen that said "+100 points!" While this feedback provided more information than was given in the no scoring rules condition, the burden was still on the player to infer the underlying cause of why points were earned or lost.

Explanation of scoring rules. Players were provided an explanation about the game's underlying scoring rules information in the (a) tutorials prior to game play, (b) general help menu that was available throughout game play, and (c) math elaborated feedback that made explicit the link between the event in the game and the mathematical principle underlying that event. The explanation given in the tutorials included which specific events triggered the earning or losing of points, as well as the rationale for *why* these events were chosen.

Explanation of scoring rules plus incentive to seek additional feedback. Players in this condition played an identical version of the game as those who were given an explanation of the scoring rules, with one exception. When the players lost points, by making a critical mistake in the game, an option to regain a portion (%) of the points if they accessed feedback was given. In the remaining text, this condition will be referred to as the scoring explanation with incentive group.

Experimental Materials: Game

Designing the game. The game developed for this study, *Save Patch*, was a collaboration among researchers at the National Center for Research, Evaluation, Standards, and Student Testing (CRESST) and graduate students in a game program at the University of Southern California's (USC) Game Innovation Lab. The expertise of the researchers at CRESST resided primarily in the areas of math, assessment, and learning; the USC students were studying game design at the Game Innovation Lab. The USC students were in their second year of school, pursuing a master of fine arts in game design and were supervised by the director for the Game Innovation Lab. A series of earlier studies of the *Save Patch* game conducted with middle school students indicated that for students with low prior knowledge, the game increased performance on a math assessment from pretest to posttest (Chung et al., 2010). Given the targeted topics, however, it was believed that the game may be better suited for younger students who may have had less exposure to instruction on fractions.

Assessment architecture for game design. To ensure alignment among learning objectives, assessment items, and game design, an assessment architecture was created to

[3] Note, in this condition, students were not explicitly told which events would trigger the earning or losing of points prior to their occurrence.

guide the design of the game (Baker & Delacruz, 2007). The assessment architecture was derived from Baker's (1997) Model-Based Assessment approach, which requires specifications of the cognitive demands and key learning content to be learned.

A sample of the underlying assessment architecture for *Save Patch* is contained in Appendix A (Delacruz, Chung, & Baker, 2010). It summarizes the interrelations between the targeted learning objectives and how they were operationalized in the game design. Moving along each column illustrates the process of identifying the cognitive demands of the learning environment (i.e., problem solving), how problem solving was defined in terms of the math content, and how these ideas were embedded in the game mechanics.

Knowledge specifications. The targeted domain of this study was rational number equivalence. Creation of the assessment architecture required identification of key foundational ideas that formed the learning objectives of instruction, assessment, and game play. This information was written in the form of verbal statements called the *knowledge specifications.* The full set of knowledge specifications that were used to develop *Save Patch* are contained in Appendix B. A more extensive discussion of how the knowledge specifications were developed is found in Vendlinski, Delacruz, Buschang, Chung, and Baker (2010). Two key mathematics ideas that emerged from the literature (Lamon, 1999; Wu, 2001) and in discussions between math educators and researchers were selected as the focal concepts of the game:

1. Only identical units can be added to create a single numerical sum.
2. The size of a rational number is relative to how one whole unit is defined.

The game design was meant to address some of the misconceptions or common errors students make when adding rational numbers (e.g., students add both the numerator and denominator of two rational number addends) (Brown & Quinn, 2006; Driscoll, 1982). One explanation for this common error is that students have difficulty understanding the meaning of a rational number (Kilpatrick, Swafford, & Findell, 2001). Specifically, students do not adequately decode the "unit" of a rational number (Fuson, 2003; Lamon, 1999). The difficulty with understanding the meaning of rational numbers is further complicated by the fact that the concept of rational numbers is often depicted as representations of circles (Saxe, Gearhart, & Seltzer, 1999). Therefore, to help make the connection to the prior knowledge students may have about whole number addition, the game exploited the fact that real numbers can be broken into smaller, identical parts to facilitate fractional addition. Specifically, the game was designed to illustrate that this process of the decomposition of numbers is similar in both integer and fractional addition. It also presents the concept of rational numbers in multiple representations (number lines, grids, and symbolic numbers).

Game play. Unlike previous games designed to teach mathematics (e.g., *MathBlaster*), fluency with the basic ideas (learning goals) was integral, not ancillary, to game play. The objective of the game was to help the game character (Patch) jump over obstacles (e.g., spikes, lava, quicksand) and move from block to block to reach the last "X" block (the final goal). To do this, players needed to compute the distance of the jump, place trampolines on the blocks, and add enough coils to the trampolines to make Patch bounce. The size of the coil determined how far Patch would bounce. For example, a one-half unit coil would cause Patch to jump over a one-half unit interval.

The first part of the game required the player to determine the size of the intervals of the grid. The intersection of vertical red bars indicated the boundaries of the whole unit. The green dots broke up the whole unit into intervals (see Figure 1).

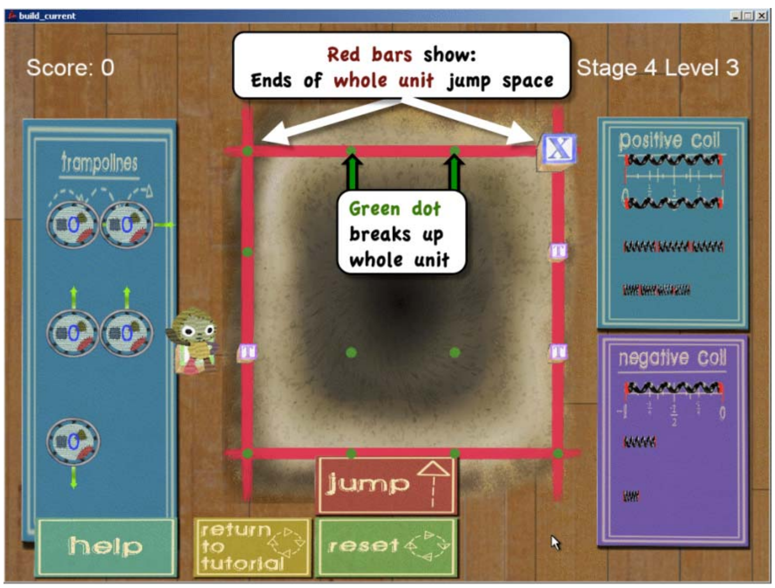


Figure 1. The intersection of the vertical bars depicted the boundaries of the whole unit. The green dots broke up the whole unit into intervals.

The size of the intervals reflected how the whole unit had been broken into fractional pieces. For example, in Figure 2, the size of the interval is one third of a unit because there are three intervals between the red bars. The same principle (size of a rational number is relative to the whole) was also applied to the sizes of the coils. In Figure 2,the size of the coil piece was one third of a whole unit coil because the whole unit coil had been divided into three pieces.

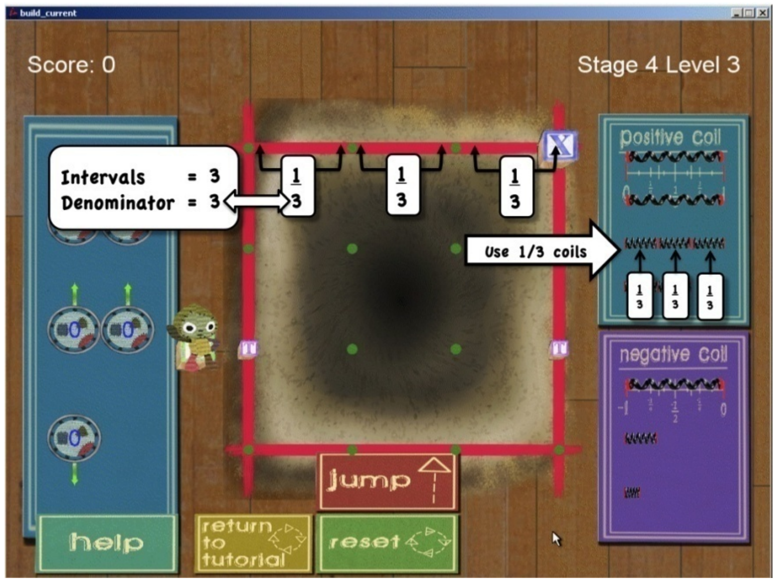


Figure 2. The number of pieces the whole unit has been broken into determined the size of the interval/fraction.

Once the players have figured out the unit size of the spaces on the grid, they needed to add together the correct number of coils that will span the distance to jump over. For example, in Figure 3, to get safely from one block to the next, Patch needs to jump over three one-third-unit intervals. This means that to successfully make it to the next block, the trampoline must contain three one-third-unit-sized coils. If the trampolines did not contain the correct number or size of coils necessary to get Patch to the next block, Patch exploded into feathers and the students were allowed to replay the level.[4] To reinforce the idea that only fractions with like denominators can be added together, the player could not combine coils that have different unit sizes. If a player had a trampoline with a one-third coil on it and tried to add a one-sixth coil, the one-sixth coil would not go onto the trampoline.

[4] There were no limits to the number of times a student could attempt a move or reset a level.

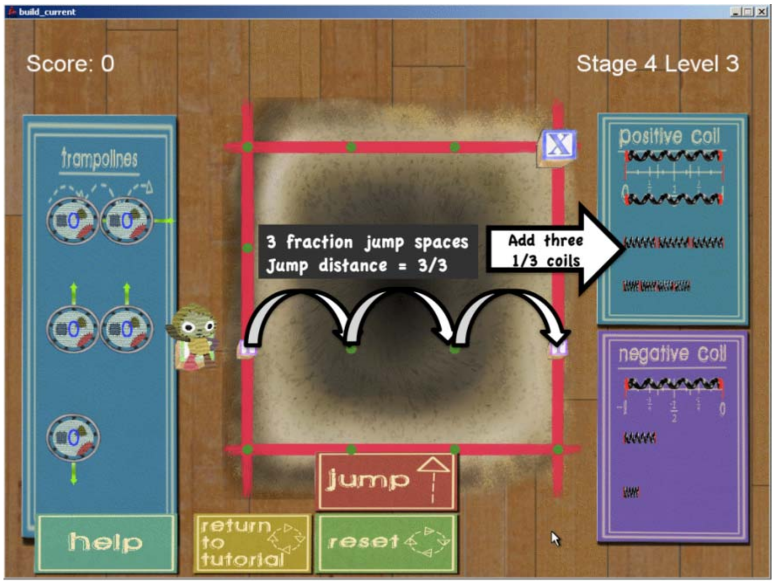


Figure 3. Jump distance equals total number of spaces between blocks. The number of coils to add to a trampoline equals the jump distance.

Scored events during game play. Across all of the conditions, three events were chosen as key points where performance would be evaluated. These events were chosen because they mapped onto the learning objectives of the game (see Appendix C for the rationale underlying the choice of the scored events). Points were *earned* any time the following event occurred: (a) player used coils that were the correct unit size for the grid, (b) player added together coils with like denominators, and (c) player successfully completed the level. Points were *lost* when any of the following events occurred: (a) player used coils that were not the correct unit size for the grid, (b) player attempted to add coils with unlike denominators together, and (c) player failed to get Patch to an intermediate goal.

Stages and levels. A stage and level approach was implemented. The game was divided into six stages that represented chunks of information that were taught in the game. Each stage began with a tutorial, followed by the game levels for student practice of what was taught in the tutorials.

Tutorials. In this section, the structure and content of the tutorials are provided. First, information about what was constant across the treatment conditions is given. Then, details of the differences in the tutorials among the treatment conditions are discussed.

Similarities across treatments. For each condition, each of the six stages began with a tutorial level, which introduced the math and game content (see Appendix D for example screenshots of the tutorial). The design of the tutorial was meant to first contextualize the math concept within the game, and then discuss how that concept relates to math in general.

For example, in Stage 2, players were introduced to the concept that *the size of a fraction is relative to the whole unit*. With the purpose of figuring out the size of the spaces on the grid, it was explained to players that the denominator of the size of the space represents the number of spaces the whole unit was broken into. Players were directed to the vertical red lines as the whole unit boundaries and the green dots that marked the intervals. Next, this same concept was illustrated in the context of determining the size of a coil. Finally, players were told that this same concept applies to determining the size of a fraction in the context of general math using symbolic, Arabic numerals.

The tutorial information was presented as both written text and guided interaction. The guided interaction took the form of asking students to perform a task such as "drag the 1/2-unit coil to the trampoline." If the task was executed correctly, the student advanced to the next part of the game. However, if the task was not correctly executed, feedback was given (e.g., "No, that is the 1/4-unit coil") and the students were given the option to try again. There were no limits to how many attempts a student could make during the guided tutorials.

The tutorials were presented as separate screens. To advance to the next screen, the student had to click on the "Next" button. Although students could not move back and forth between the screens, the option to return to the beginning of the tutorial (both during the tutorial itself and during the practice game levels) was always available. The content of the stage determined the number of screens presented (see Table 2 for a brief description of what is taught at each stage and the number of screens per stage). On average, students spent between 1 to 6 minutes on the tutorials, depending on the stage.

Table 2

Stage and Level Description for the Game

Stage	Description	Number of screens presented
1	Basic game mechanics: Using trampolines and coils	12
2	Defining a unit, identifying the fraction size, meaning of a denominator	8
3	Adding fractions with like denominators, meaning of a numerator	5
4	Moving in different directions, using negative coils (to "undo" the addition of coils)	2
5	Decomposing a quantity into smaller, equal parts using "scrolling"	3
6	Adding fractions with unlike denominators	5
Bonus	More complex adding of fractions with unlike denominators	8

Tutorial information differences among treatments. The only differences in the content of the tutorials among the treatments pertained to the game's scoring rules. All of the students in each condition were told where the points were located on the screen. However, only students in the scoring explanation and scoring rules with incentive groups were provided information during the tutorials on which events were scored and the underlying mathematical reason. For instance, in Stage 3, players in these two conditions were told when points would be earned (added coils with like denominators) and when points were lost (added coils with unlike denominators). Players in these two conditions were also given the rationale for why points were lost when coils with unlike denominators were added together (see Figure 4).

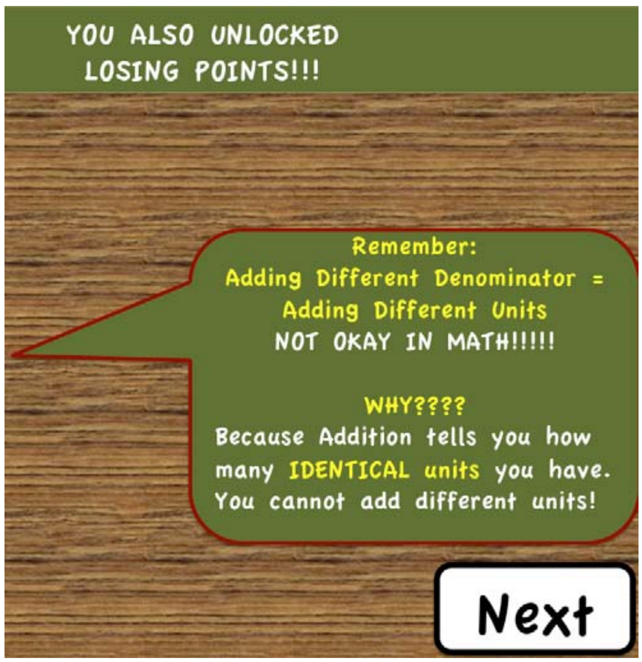


Figure 4 .Example of information given to player about the rationale underlying a scored event.

Feedback. All of the players in each condition received feedback that was in the form of the knowledge of results without elaboration or explanation (Shute, 2008). This type of feedback was given after three events. When a player tried to add a coil to a trampoline that had a coil with unlike denominators, the coils would not combine on the trampoline. Also, if the trampoline did not contain the correct-sized coil, or amount necessary to cover the distance of the jump, Patch exploded into feathers. When Patch jumped from block to block without exploding, this indicated that the player placed the correct quantity of coils needed to make the jump. The score displayed on the screen also changed after each level was completed.

Feedback differences among conditions. What differed among the conditions was the amount and type of explanatory scoring-based feedback given. As stated before, students in the points-only feedback, scoring explanation, and scoring explanation with incentive groups were explicitly alerted when points were earned or lost.

Players in the scoring explanation and scoring explanation with incentive groups were given a feedback image intended to serve multiple functions. The feedback image (a) indicated that points were earned or lost, (b) gave each event a mathematical label (e.g., correct denominator bonus), and (c) made explicit the link with the underlying mathematics and the game event (e.g., good coil choice) (see Appendix E for screenshots of each feedback image).

General help menu. All of the players in each condition had access to a general help menu at any time throughout the game levels.[5] The topics that were included in the general help menu provided information on game mechanics (e.g., how to make Patch move in different directions) as well as instructional information such as how to choose the right-sized coil and how to add coils of different sizes. The game kept track of which topics were accessed and how often they were accessed.

Like the tutorials, the only content in the general help menu that differed among the conditions was information pertaining to the scoring rules. Players in the scoring explanation and scoring explanation with incentive groups had access to an additional topic labeled "Earning points." This information was a summary of the scoring rules information that was presented in the tutorials.

Feedback help. After the second consecutive mistake (e.g., chose the wrong-sized coil, added coils with different denominators, failed to reach the goal), all of the players in each condition were given an opportunity to access additional feedback by clicking on a button that read, "Click here for help." The additional feedback provided elaborated explanations and hints that were designed to assist the player with repairing the mistake. Appendix F contains example screenshots of the additional feedback that was given in the game.

Because of a programming error, one of the treatment conditions (no scoring rules information) was not given an opportunity to access instructional feedback help after a failed attempt. However, it is believed that any potential problems were somewhat mitigated by the fact that the same information was made available in the general help menu.[6] Additionally, in

[5] The general help menu was not accessible during the tutorials.
[6] The data indicate that 24 out of the 28 students in the no scoring scheme information condition accessed the same information at least once. Eighteen students accessed it more than once.

the analyses of the data, the data from this condition were not examined individually (i.e., they were combined with the data from students in the points-only feedback). The rationale for this decision is provided in the results section.

Materials: Dependent Measures

Math achievement measures. The pretest and posttest were designed to assess the targeted knowledge specifications of the game. Although the knowledge specifications identified what concepts were to be assessed, *item specifications* specified the performance expectations of a student proficient with the knowledge (Vendlinski et al., 2010). Specifically, the item specifications identified the stimuli or situations that could be presented to the students and the subsequent desired responses (see Appendix G for item specifications).

The assessment used to measure math achievement consisted of items that focused on the size of a unit and how to represent and add fractions with both like and unlike denominators. See Figure 5 for an example of the assessment items that were administered. The full measures are in Appendix H (pretest) and Appendix I (posttest).

What does the denominator of 4 tell you in $\frac{3}{4}$? **Choose only one answer.**

a. It tells you there are four $\frac{3}{4}$'s in this fraction
b. It tells you the whole unit is broken into four pieces
c. It tells you there are four whole units in this fraction
d. It tells you to add 3 four times

Figure 5 .Sample items on the math assessment.

There were 31 items on the pretest and the same 31 items also appeared on the posttest. The posttest contained an additional 13 items. These additional items targeted the same concepts as the items that appeared on both the pretest and posttest. However, these items were consistent with practice in the game (for example, see Figure 6).

Suppose a trampoline with a $\frac{3}{2}$ coil was placed on the grid as shown below.

Place an "X" on the spot where Patch would land after pressing *Jump*.

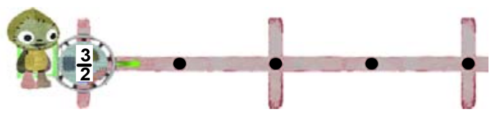


Figure 6. Example of a posttest item that was written in the game context.

Game performance outcomes. The game outcomes described in Table 3 were identified prior to game development as critical indices of learning. These events were automatically logged during game play to facilitate analysis of in-game learning.

Table 3

Game Performance Outcome Descriptions

Description
Total number of coils with unlike denominators added together
Total number of coils that were the wrong-sized unit for the grid was used
Total number of coils with like denominators added together
Maximum level reached after 20 minutes
Total number of resets in the game
Total number of failed attempts in the game

Frequency of voluntarily accessing additional feedback. The frequency of voluntarily accessing additional feedback was measured by examining the proportion of times feedback was accessed, relative to the number of opportunities to access feedback. Four measures were created by computing the number of times feedback was accessed divided by the number of times an opportunity to access the feedback was provided. These were: the total proportion of times that the additional feedback was accessed (a) overall in the game, (b) after coils with different denominators were added together, (c) when a coil that was the wrong-sized unit for the grid was used, and (d) when Patch did not make it to the next block.

Post-Game Survey: Motivation and Student Experience with Games and Mathematics

A survey was administered after the game was played. The questions in the survey gathered self-reported motivation to play the game and information about a student's computer and gaming experience. Additional measures were obtained for variables that might be related to games for learning.

For example, research has shown that achievement can be predicted by two general components: value and expectancy (Pintrich & De Groot, 1990; Wigfield & Eccles, 2000). It is also argued that self-efficacy (i.e., confidence about success) is an integral part of gaming (Garris, Ahlers, & Driskell, 2002). Additionally, part of the appeal of games for learning is that students can engage in both collaborative and competitive play (Kirriemuir & McFarlane, 2004), which research has shown can produce academic benefits (Webb & Palinscar, 1996). Therefore, the post-game survey contained measures of academic self-efficacy, math self-concept, and preferences for competitive or cooperative learning environments. The full measure can be found in Appendix J.

Self-report on motivation. Students rated how much they agreed with statements about the game, such as “I lost track of time” or “I thought the game was boring” on a four-point scale (1 = *disagree*, 2 = *disagree a little*, 3 = *agree a little*, 4 = *agree*).

Computer and gaming experience, Students reported how frequently they played nine types of games (e.g., puzzle, real time strategy, action) on a four-point scale (1 = *hardly ever*, 2 = *sometimes*, 3 = *often*, 4 = *very often*). Students also reported how many hours per week they played video games, as well as their skill level with video games and computers.

Self-efficacy and math self-concept. For the measures of self-efficacy and math self-concept, four of the scales from the Programme for International Student Assessment (PISA) were used (Marsh, Hau, Artelt, Baumert, & Peschar, 2006). For both measures, students rated how much they agreed with the given statements on a four-point scale (1 = *disagree*, 2 = *disagree a little*, 3 = *agree a little*, 4 = *agree*). The reliability information on these scales reported by Marsh et al. (2006) is presented in Table 4.

Table 4

Reliability Information on PISA Scales for Measures of Self-Efficacy and Math Self-Concept

Current study measure	PISA scale	Number of items	Cronbach’s alpha for total sample
Self-efficacy: 8 items total	Perceived self-efficacy	4	.77
	Control expectation	4	.75
Math self-concept: 6 items total	Interest in math	3	.75
	Self-concept in mathematics	3	.88

Cooperative and competitive learning preferences. The items from PISA that targeted competitive learning and cooperative learning preferences were used to measure students' preferences for cooperative and competitive learning. Students rated how much they agreed with the given statements on a four-point scale (1 = disagree, 2 = disagree a little, 3 = agree a little, 4 = agree). Table 5 contains the reliability information reported by Marsh et al. (2006) for these scales.

Table 5

Reliability Information on PISA Scales for Measures of Cooperative and Competitive Learning Preferences

Current study measure	PISA scale	Number of items	Cronbach's alpha for total sample
Competitive learning preference	Competitive learning	4	.78
Cooperative learning preference	Cooperative learning	5	.75

Procedures

Recruitment of Schools and Participants

Description of the after-school program. Schools and participants were recruited from an after-school program (from fourth to sixth grades) in a large urban city that was run by the city's Department of Parks and Recreation. The program coordinated the after-school activities at 62 schools in the city. At each school, there was a site director and a team of program leaders that ran the program. The components of the after-school curriculum included academic enrichment through homework and tutoring assistance, literacy and math activities, and enrichment recreation activities, such as art, cooking, and chess. The majority of the students that participated in the program matched the targeted sample of this study (i.e., groups that historically do poorly in math). A large proportion of the students were either African American or Latino/a, were eligible for a free lunch, and were classified as English language learners (ELL) (Fitzgerald, 2009). Five sites were chosen on the basis of their proximity to each other to facilitate data collection.

Participants

The data were collected from 161 students in a large urban city, at five different elementary schools. Table 6 contains information about the schools used in the study. The schools serve a largely socioeconomically disadvantaged population (low SES), with a sizable proportion of ELL students. Moreover, other than School A, most of the schools that participated in the study did not meet the necessary proficiency levels in mathematics. However, compared to schools with comparable students, their performance was generally above average.

Table 6

Information About Schools Used in the Study

School	Number of students (4th-6th grade)	Percentage of students classified as: Low SES	English language learner	Base API in 2009	API rank[a]	Met Adequate Yearly Progress in 2008-09 for math proficiency	Percentage of students that achieved proficient or advanced level on state standardized tests: Low SES	English language learner
School A	173	100%	35%	752	8	Yes	55%	46%
School B	274	100%	52%	702	4	No	39%	30%
School C	358	70%	31%	716	6	No	37%	29%
School D	130	90%	14%	630	5	No	31%	27%
School E	242	98%	44%	747	7	No	54%	39%

[a]Similar schools.

Student background. The students were in fourth to sixth grade. All of the participants in this study participated in the district's after-school program.

Background information was collected from students in a survey administered after they played the game. The numbers in the table below do not reflect the entire sample because many students were unable to complete the survey due to early departures. This information has not been otherwise verified.

Table 7

Demographic Information: Grades, English Language Learner Status, and Self-Reported Math Grades

Background variable	n (%)
Grade in school	
Grade 4	39 (30.5%)
Grade 5	42 (32.8%)
Grade 6	30 (23.4%)
Gender	
Female	51 (39.8%)
Male	53 (41.4%)
Ethnicity	
Biracial/multiethnic	13 (10.2%)
African American	28 (20.1%)
Asian or Pacific Islander	17 (13.3%)
Latino	31 (24.2%)
Native American	2 (1.6%)
White, non-Hispanic	4 (3.1%)
Other	13 (10.2%)
English Language Learners: How often do people in your home talk to each other in a language other than English?	
Never	32 (25.0%)
Once in a while	16 (12.5%)
About half of the time	17 (13.3%)
All or most of the time	40 (31.3%)
Self-reported prior math grades on last report card	
A	43 (33.6%)
B	24 (18.8%)
C	10 (7.8%)
D	3 (2.3%)
F	1 (.8%)

Data Collection Administration Procedures

This section will describe the procedures that were used during the study administration. First, it will address how students were assigned to a treatment condition. Then it will outline the basic setup of the technology (e.g., labeling computers with numbers, entering IDs) as well as measures. Next, the study will provide a description of the basic overall administration of activities. Finally, given the unique layout of each school, the configuration of the equipment and measures at each school will be described.

Assignment to treatment condition. Within each school, students were randomly assigned to a treatment condition. The sample was not stratified by grade level. Cards were created with the numbers 1 to 35. Upon arrival to the data collection site, the student was given a card with a number. This process was used to ensure random assignment to each condition. However, when there was an uneven number of students at a site, priority was given to assigning students to one of the treatment conditions to ensure that the sample sizes for each treatment condition were sufficient. For instance, when there were 18 students in a group, after the 15 students were distributed across the five conditions, the remaining three students were assigned to the *Save Patch* treatment conditions (rather than the control group). This process contributed to there being fewer students in the control group. After the students received their numbers, they sat down at their assigned space. All of the students were told to wait until they were told to begin.

Setup of laptops and tests. Thirty laptops and five iPod Touches were available at each administration. The four versions of the game were on each laptop. Each laptop and iPod Touch was assigned a data collection computer number (laptops: 1 to 30, iPod Touches: 1 to 5).These numbers were printed on large cards that were placed on or near the computers.

Also, the pretest was labeled with an ID and placed at each laptop prior to the study; the participant's ID was entered onto each computer and iPod Touch. The computers and iPod Touches were organized so that students within each condition were in close proximity to each other.

Data collection activities. An average of about 90 minutes was spent on the study at each site. The basic timing and order of the task were the same at each site. The descriptions of the study activities will follow.

Brief introductory announcement (5 min). During the introductory announcement, the researchers introduced themselves and provided background information regarding the purpose of the study. Students were told that there were different versions of the game and that the purpose of the study was to investigate which version was most effective for learning.

Pretest (15 min).When students finished the pretest, they were asked to wait to begin to play the game until everyone finished the test. Students were told that the pretest was not going to count for a grade; instead, it would be used to determine whether the game was effective for learning math. Also, the students were instructed to write “don’t know” or “DK” if they did not know the answer to a question, rather than leave it blank. Students began the pretest at the same time. On average, it took students about 10 to 15 minutes to complete the test, although the time necessary to complete the test never exceeded 15 minutes.

Game introduction (2 min). Once everyone finished the pretest, the game was introduced. Students were told that there was very helpful information given in the game, and that it was important to read the information because there was no audio. Students were also made aware of the help menu that was available throughout the game.

Game play (25-40 min). Every effort was made to ensure that each student played the game for at least 30 minutes. However, the amount of time for game play varied because many students departed for home early.

Posttest and survey (30 min). After playing the game, the students were given the posttest and the demographic/motivation survey. Students were told that some of the questions that appeared on the pretest would be on the posttest, but that the information they learned from playing the game might be useful to answer the posttest questions.

School-specific configurations. The data collection configuration varied at each school. The next section will describe the configuration for each of the schools and how the study was administered at each site.

School A. Data were collected from 44 students at this site. Because of the number of available laptops and iPod Touches, the students were randomly split into two groups and the study took place in two overlapping cycles (see Table 8). There were 26 students in the first group and 18 students in the second group. The study took place in two classrooms, with one room devoted to game play (Room A) and the other used to complete some of the paper measures (Room B). The students moved from classroom to classroom to maximize space and time on the computer. Group 1 took the pretest and played the game in Room A; they took the posttest in Room B. Group 2 completed the pretest in Room B and played the game and took the posttest in Room A. When Group 1 began their game play, Group 2 began their pretest in Room B.

Table 8

Data Collection Administration Timing at School A

Room A	Time	Room B	Time
Group 1: Pretest	15 min.		
Group 1: Game play	40 min.	Group 2: Pretest	15 min.
Group 2: Game play	30 min.	Group 1: Posttest	30 min.
Group 2: Posttest	30 min.		

School B. Data were collected from 27 students at this site. The study took place in one classroom. There was additional time after the posttests were completed, so the students were given the opportunity to continue to play if they desired. Most of the students elected to keep playing the game rather than leave the room to engage in free-choice activities.

School C. Data were collected from 32 students at this site. The study took place in the school's learning center, which was configured as both a library and a computer lab. There were six pods of four to six students throughout the room, and one big table in the middle that held 10 students. Because of the limited number of laptops, the data were collected from two cycles of groups at this site. For the first cycle, there were 23 students who were distributed across three pods and the big table. The students in the first cycle who were seated at the big table moved to the smaller pods where they completed their posttests. The second cycle of students were moved to the big table and completed their study.

School D. Data were collected from 24 students at this site. The study took place in one classroom.

School E. Data were collected from 34 students at this site. The study took place in the cafeteria. Seven long tables were used (i.e., three of the tables were used as the computer area; three of the tables were used to complete the paper measures; and one table was used for the students in the iPod Touch condition). The sample was divided into two groups for convenience. The first group moved between the computer and paper measure areas. The second group stayed at the same space for the study. Data were collected from two groups with 30 students in the first group and four students in the second group. Each of the four students in the second group was randomly assigned to one of the four treatments.

Analysis

The first phase of analysis consisted of determining the final sample of students. Only students who completed both the pretest and posttest, and played the game for at least 20 minutes were included in the final analyses. The cause for attrition was that data collection took place during an after-school program. Many students were unable to complete the study due to early departures. One student opted not to complete the study and was in the no scoring rules information condition.

Table 9

Breakdown of Students from Original Sample and Final Sample

Sample of students	Control group	No scoring rules information	Points-only feedback	Explanation of scoring rules	Explanation of scoring rules with added incentive	Total number of students
Students in original sample	16	34	32	41	39	162
Students in final sample	16	28	22	30	32	128

Analysis of Dependent Measures

Math achievement measures. The second phase of analyses involved scoring the items on the math assessment. The three open-ended items on the math assessment were scored using three raters who were trained on how to apply the rubrics. The final rubrics that were used to score the items are contained in Appendix K. The training process began with each rater being given five student responses to each of the three open-ended questions. Ratings were reviewed as a group and discrepancies were resolved. The rubric was modified for clarity as needed. For each of the three questions, each rater used the final version of the rubric to score responses from 20 students (10% of sample). Interrater reliability for each item was determined by calculating the Intraclass Correlation Coefficients (ICC) for each item (McGraw & Wong, 1998). A two-way, mixed-effect model was used to examine the absolute agreement of measurements between the three raters for each of the three open-ended items (see Appendix L for the rationale underlying the choice of the model used). Table 10 contains the interrater reliability for the three open-ended items.

Table 10

Interrater Reliability for Open-Ended Items

Item	Interrater reliability
1	0.75
2	1.00
3	1.00

After the open-ended items were scored, four math achievement measures were created from the assessment items: (a) pretest, (b) posttest, (c), transfer items, and (d) game context items. Table 11 provides a summary of these math achievement measures.

Table 11

Description of Math Achievement Measures

Math achievement measure	Description	Total number of items
Pretest score	Total number of correct responses on the pretest	31
Posttest score	Total number of correct responses on the posttest items that appeared on the pretest	31
Transfer items score	Items that asked students to represent fractions with different representations (e.g., cookies) or word problems	4
Game context items score	Items that were written in the same context of the game and directly reflected what was practiced in the game	13

Normalized change scores. Hake (1998) proposed the use of normalized gain scores (see Equation 1) as a way to better reflect the strength of the treatment, giving more weight to students who did better on the pretest. The rationale for the use of normalized gain scores is that it is more difficult to improve performance for students who are already doing well in a domain (i.e., the ceiling effect). The use of normalized gain scores, or *g*, is most often used in Physics education.

Equation (1): $$g = \frac{posttest(\%) - pretest(\%)}{100 - pretest(\%)}$$

However, Marx and Cummings (2007) presented two problems with using *g* to measure change that are relevant to this study. First, the authors claimed that *g* has a bias toward low pretest scores and that *g* generates a non-symmetric range of scores, which makes finding the average *g* of a group difficult. To correct for these problems, the use of a normalized *change* score (denoted as *c*) is proposed (see Equation 2).The normalized change score characterizes the change in scores regardless of the pretest score and eliminates the low-pretest-score bias, with normalized change scores ranging from -1 to +1. Moreover, students whose posttest scores either increase or decrease by the same percentage relative to the maximum possible gain or maximum loss will obtain the same magnitude of *c*. Because there were students for whom pretest to posttest scores decreased in this current study, it was logical to use an analytical tool that more appropriately accounts for the decrease in scores.

$$\text{Equation (2): } c = \begin{cases} \dfrac{posttest(\%) - pretest(\%)}{100 - pretest(\%)} & \text{pretest} > \text{posttest} \\ 0 & \text{posttest} = \text{pretest} \\ \dfrac{posttest(\%) - pretest(\%)}{pretest(\%)} & \text{posttest} > \text{pretest} \end{cases}$$

Extracting Outcomes From Log Data: Game Performance and Help Seeking

The second phase of the study was to extract game performance and help-seeking outcomes from the log data. The log data from the games were downloaded from each computer and compiled into a single text file. The text file was then exported into SPSS for processing. Creation of the game performance and help-seeking outcomes required basic summation or averages of the event-level data (e.g., total number of resets).

Measures of the Motivation to Play the Game

Two measures of students' self-reported motivation to play the game were created from the mean average of responses to the survey items. The first measure comprised all of the positive comments about the game (10 items) and the second measure consisted of all of the negative comments about the game (8 items).

Reliability Analyses

To determine the inter-item reliability of the scales that were created, Cronbach's alpha were computed for the (a) pretest, (b) posttest, (c) transfer items, (d) game context items, (e) motivation measures, (f) measures of self-efficacy and math self-concept, and (g) measures of preferences for cooperative and competitive learning.

Analyses of Research Questions

For each of the research questions, descriptive statistics were obtained from the data by computing the means, standard deviations, and standard error. Pearson's correlation coefficients were computed to examine the correlation among the measures or variables.

For the first three research questions, analyses of variances (ANOVAs) were computed to test the effect of the treatment conditions on the math achievement measures and the game performance measures. Linear regressions models were used to examine the additive effect of treatment conditions after adjusting for pretest scores. Interactions between pretest scores and treatment conditions on the math achievement measures were also tested.

Hypothesis testing. Three sets of orthogonal planned comparisons were conducted to examine if the data supported the hypotheses of the study. Performance on the math achievement measures, game performance measures, and the frequency of accessing feedback were compared between students in the:

1. Scoring explanation and minimal scoring information conditions (Hypothesis 1)
2. Scoring explanation plus incentive and the scoring explanation conditions (Hypothesis 2)
3. Scoring explanation plus incentive and minimal information conditions (Hypothesis 3)

Exploratory Analyses

Interaction effects of background variables on math achievement were measured. ANOVAs were employed (using treatment conditions as the independent variable, math achievement measures as the dependent measures, and the background variables that were correlated with the math achievement measures as covariates). Since the study was administered at five different schools, analyses were conducted to rule out any potential school-specific effects. Analyses of subsamples were also conducted to examine if the treatment conditions had differential effects on students with certain characteristics (e.g., low self-efficacy).

Closer examination of access of feedback in individual levels. To examine more closely the differential effects of the treatment conditions on the access of feedback after the wrong-sized unit was used, the data were examined and levels were identified in which all three of the treatment conditions accessed the feedback. Three measures were created: (1) the amount of time it took to access the feedback, (2) the amount of time that was spent on the feedback, and (3) the amount of time it took to complete the level after feedback was accessed.

Access of the general help menu. As an additional analysis of the effect of the treatment conditions on accessing feedback in general, the frequency of accessing each topic in the general help menu was determined by each condition.

Motivation. To determine whether the treatment conditions varied in their self-reported motivation to play the game, ANOVAs were computed. The independent variable was treatment condition and dependent measures were two scales that separately comprised (a) the positive comments and (b) the negative comments.

Results

The following section contains the results from the analyses of the data. First, the inter-item reliability for each of the scales that were created for the study is presented. Second, the first research question is addressed by reporting results on the effectiveness of combined *Save Patch* treatments compared to a control. Third, the results from the verification of the randomization of assignment to treatment conditions are given. Descriptive, correlation, and inferential analyses of the effects of the treatment conditions on the math achievement measures, game performance measures, and the voluntary access of feedback are reported. Results from the additional exploratory analyses are presented: (a) tests of interactions to discern differential treatment effects on subgroups classified according to background variables, (b) analyses of voluntarily accessing feedback in specific levels, and (c) comparisons between the treatment groups on the access of topics in the general menu.

Reliability Analyses

Cronbach's alpha coefficients were computed to determine the inter-item reliability of each scale, reported in Table 12. The transfer item scale yielded a low reliability coefficient. The low reliability suggests that this scale may not be unidimensional, which may make interpretations using the transfer measure questionable. However, to examine further the reliability of the transfer measure, the test-retest reliability was determined by computing a correlation between the transfer item scales on the pretest and on the posttest for the control group. The reliability coefficient for the transfer items was .83.

Table 12

Inter-item Reliability Results for Scales Used in Study

Measure	Number of items	Cronbach's alpha coefficient
Pretest	31	0.89
Posttest	31	0.87
Transfer items	4	0.33
Game context items	13	0.83
Motivation: Positive comments	10	0.92
Motivation: Negative comments	8	0.80
Self-efficacy	8	0.88
Math self-concept	6	0.82
Cooperative learning preferences	5	0.82
Competitive learning preferences	4	0.73

Research Question 1. What is the effect of a pre-algebra math game on learning outcomes of upper elementary students compared to a control group?

The first research question asked whether there would be an effect of playing *Save Patch* compared to a control group on the math achievement measures[7] including the posttest, normalized gain score, and transfer items. Table 13 reports descriptive information about the two groups, including means, standard deviations, and standard error. The control group's average normalized change score was negative.

Table 13

Descriptive Statistics for the Control Group and Combined *Save Patch* Treatments

Math achievement measures	Min.	Max.	*M*	*SE*	*SD*
Control group[a]					
Pretest	6.00	26.00	14.44	1.35	5.40
Posttest	0.00	25.00	13.81	1.71	6.85
Transfer items (pretest)	0.00	4.00	2.19	0.29	1.17
Transfer items (posttest)	0.00	4.00	2.06	0.32	1.29
Normalized change scores	-1.00	0.38	-0.08	0.10	0.39

[7] The game context items were not analyzed because the control group did not answer those items.

Math achievement measures	Min.	Max.	*M*	*SE*	*SD*
Combined *Save Patch* treatments[b]					
Pretest	2.00	31.00	15.17	0.62	6.56
Posttest	3.00	30.00	15.38	0.56	5.94
Transfer items (pretest)	0.00	4.00	2.70	0.09	0.98
Transfer items (posttest)	0.00	4.00	2.56	0.09	0.93
Normalized change scores	0.00	4.00	2.56	0.09	0.93

[a]n = 16. [b]n = 112.

Comparison of performance on math achievement measures. To compare performance on the math achievement measures, three independent samples *t* tests were conducted. The independent variable was game condition with two levels: control group and the combined *Save Patch* treatments. The dependent variables were the posttest, transfer items, and the normalized change scores. The results were not significant (reported in Table 14).

Table 14

Results from Independent Samples *t* Tests Comparing Performance on the Math Achievement Measures for Students in the Control Group and Combined *Save Patch* Treatments

						95% confidence interval of the difference		
Math achievement measures	*t*	*df*	*p*	*MD*	*SE* difference	Lower	Upper	Cohen's *d*
Posttest	-0.97	126	0.33	-1.57	1.62	-4.78	1.63	0.28
Transfer items	-1.91	126	0.06	-0.50	0.26	-1.02	0.02	0.52
Normalized change scores	-1.83	126	0.07	-0.10	0.06	-0.21	0.01	0.49

Adjusting for pretest scores as a covariate. The results just described imply that there was no effect of playing *Save Patch* on the math achievement measures. However, findings from earlier studies on *Save Patch* indicate that pretest scores tend to explain most of the variance on math achievement measures (Delacruz et al., 2010). Controlling for pretest scores may clarify the amount of learning to be attributed to playing *Save Patch.*

A linear regression framework was used to compare the differences between students in the two game conditions (control group versus playing *Save Patch*) on posttest scores and transfer item scores, after adjusting for pretest scores. The model that included both pretest scores and game conditions as predictors of posttest scores was significant, $F(2, 125) = 191.53$, $p < .001$. However, the coefficient for game condition ($\beta = .97$) was not significant, $t(125) = 1.20$, $p = .23$, which indicates that there was no effect of playing *Save Patch*, even after adjusting for pretest scores.

For the transfer item scores, the model that included both pretest scores and game conditions as predictor variables was significant, $F(2, 125) = 26.44$, $p < .001$, explaining 29.7% of the variance. The coefficient for the game condition ($\beta = .44$) was also significant, $t(125) = 1.98$, $p = .05$, with playing *Save Patch* explaining 2.2% of the variance above and beyond pretest scores, holding all other variables constant.

Data and methodology to examine the effect of the *Save Patch* treatments on math achievement measures. To answer research questions 2a and 3a, the data were examined to determine if there were any significant differences between the groups on gender, grade in school, gaming experience, self-efficacy, math self-concept, preferences for cooperative or competitive learning, and prior knowledge. The distribution of males and females did not differ for any of the *Save Patch* treatment groups, $\chi^2(3) = 3.75$, $p = .29$, nor did the grade in school of the student, $\chi^2(6) = 2.20$, $p = .90$. The ANOVAs that tested the effect of treatment conditions on students' gaming experience, self-efficacy, math self-concept, preferences for cooperative or competitive learning, and prior knowledge failed to yield significant *F* statistics for any of the variables (see Appendix M). The findings from these analyses verify that the random assignment to treatment condition effectively created equivalent groups in terms of these variables.

Dependent measures: Math achievement measures. Descriptive information for each for the math achievement measures for the overall sample is reported in Table 15.

Table 15

Descriptive Statistics for the Math Achievement Measures

Math achievement measures	Min.	Max.	*M*	*SE*	*SD*	Skewness
Pretest[a]	2.00	31.00	15.17	0.62	6.56	0.27
Posttest	3.00	30.00	15.38	0.56	5.94	0.27
Transfer items (pretest)	0.00	4.00	2.70	0.09	0.98	-0.51
Transfer items (posttest)	0.00	4.00	2.56	0.09	0.93	-0.32
Game context items	0.00	13.00	3.91	0.28	2.99	0.78
Normalized change score	-0.50	0.55	0.02	0.02	0.17	0.30

[a]$n = 112$.

Fisher's skewness coefficient for the game context items suggested some skewness to the left. However, to test the effect of the *Save Patch* treatment conditions on the math achievement measures, analyses of variances (ANOVAs) were computed, which tended to be fairly robust to violations of normality (Lindman, 1974). Next, descriptive statistics for each of the math measures are presented for each individual treatment condition in Table 16.

Table 16

Descriptive Statistics for the Math Achievement Measures by Each Treatment Condition

Math achievement measures	Min.	Max.	*M*	*SE*	*SD*	Skewness
Pretest scores						
No scoring rules information[a]	4	28	15.14	1.21	6.40	-0.03
Points-only feedback[b]	4	29	15.95	1.55	7.27	-0.09
Explanation of scoring rules[c]	4	31	15.33	1.21	6.60	0.80
Explanation of scoring rules plus incentive[d]	2	29	14.50	1.13	6.39	0.31
Posttest scores						
No scoring rules information	4	29	14.68	1.09	5.79	0.42
Points-only feedback	3	26	15.68	1.38	6.47	-0.39
Explanation of scoring rules	4	30	15.67	1.18	6.47	0.53
Explanation of scoring rules plus incentive	6	27	15.53	0.96	5.41	0.41
Transfer item scores (pretest)						
No scoring rules information	1	4	2.89	0.17	0.88	-0.14
Points-only feedback	1	4	2.86	0.21	0.99	-0.35
Explanation of scoring rules	1	4	2.60	0.17	0.93	-0.18
Explanation of scoring rules plus incentive	0	4	2.50	0.20	1.11	-0.84
Transfer item scores (posttest)						
No scoring rules information	1	4	2.57	0.16	0.84	0.17
Points-only feedback	1	4	2.86	0.19	0.89	-0.61
Explanation of scoring rules	1	4	2.53	0.18	1.01	0.01
Explanation of scoring rules plus incentive	0	4	2.38	0.17	0.94	-0.85

Math achievement measures	Min.	Max.	*M*	*SE*	*SD*	Skewness
Game context item scores (posttest)						
No scoring rules information	0	11	3.36	0.56	2.97	1.10
Points-only feedback	0	11	4.23	0.63	2.98	0.58
Explanation of scoring rules	0	13	4.33	0.65	3.58	0.84
Explanation of scoring rules plus incentive	0	9	3.78	0.43	2.42	0.10
Normalized change scores						
No scoring rules information	-0.36	0.33	-0.02	0.03	0.18	0.27
Points-only feedback	-0.25	0.38	-0.02	0.03	0.15	0.75
Explanation of scoring rules	-0.5	0.42	0.03	0.04	0.20	-0.13
Explanation of scoring rules plus incentive	-0.15	0.55	0.06	0.03	0.15	1.21

$^{a}n = 28$. $^{b}n = 22$. $^{c}n = 28$. $^{d}n = 32$.

The normalized change scores were positive for each of the two groups that received the scoring explanation, which indicated an increase in scores between the pretest and posttest. In contrast, the normalized change scores were negative for the groups that received points-only feedback, and no scoring rules information.

Correlation among math achievement measures. Next, Pearson's correlation coefficients were computed to examine the correlation among the math achievement measures. The results are reported in the correlation matrix in Table 17. Pretest scores were significantly correlated with the posttest, transfer, and game context item scores.

Table 17

Correlation Matrix Among the Math Achievement Measures

Math achievement measures	Pretest	Posttest	Transfer items	Game context items	Normalized change scores
Pretest	1.00	0.86**	0.52**	0.54**	-0.15
Posttest		1.00	0.61**	0.59**	0.36**
Transfer items			1.00	0.34**	0.26**
Game context items				1.00	0.10
Normalized change scores					1.00

**Correlation is significant at the 0.01 level (two-tailed).

Rationale for pooling the points-only and no scoring rules information data. The purpose of pooling the data from the points-only and no scoring rules conditions was to create a condition that represented a minimal scoring information group. Two independent samples *t* tests were computed to compare the groups on the math achievement measures to ensure they did not differ significantly. There were no significant differences between the two groups on the posttest, $t(48) = .58$, $p = .57$, transfer items, $t(48) = 1.19$, $p = .24$, or the game context item scores, $t(48) = 1.03$, $p = .31$. Therefore, for the purposes of testing the three hypotheses of the study, the data from students in the points-only and no scoring rules groups were pooled together and henceforth referred to as the minimal scoring information group.

Effect of the *Save Patch* Treatment on Math Achievement

The next set of reported results are from the inferential analyses that were used to answer the following two research questions:

> Research Question 2a: What is the effect of providing an explanation about the scoring rules on math learning?
>
> Research Question 3a: Does providing an incentive to seek additional feedback affect math learning?

First, results from the ANOVAs computed to examine the effect of treatment condition on the math achievement measures are presented in Table 18. There was no significant effect of treatment scores on any of the measures.

Table 18

Results from the ANOVAs for the Math Achievement Measures: Posttest, Game Context Items, and Normalized Change Score

Source	Dependent variable	*SS*	*df*	*MS*	*F*	*p*	Partial Eta^2
Save Patch treatment conditions	Posttest	6.57	2	3.29	0.09	0.91	0.02
	Transfer items	737.53	3	245.84	286.70	0.00	0.89
	Game context items	7.35	2	3.67	0.41	0.56	0.07
	Normalized change score	0.13	2	0.07	0.10	0.10	0.41
Error	Posttest	3345.19	109	30.69			
	Game context items	987.76	109	9.06			
	Normalized change score	2.48	109	0.02			
Total	Posttest	27393.00	112				
	Game context items	2708.00	112				
	Normalized change score	2.61	112				

Differential effects of the treatment conditions on math achievement measures. Next, the results from the orthogonal planned comparisons to confirm the three hypotheses pertaining to the math achievement measures will be presented (see Table 19). The hypotheses are presented prior to giving the results of the analyses and each of the hypotheses are discussed individually.

Table 19

Results from the Planned Comparisons of Math Achievement

Hypothesis tested	Math measure	*MD*	*SE*	*t*	*p*	Cohen's *d*
Scoring explanation > Minimal scoring	Posttest items	0.49	1.39	0.35	0.73	0.09
	Transfer items	-0.17	0.22	0.78	0.44	0.02
	Game context items	0.54	0.70	0.78	0.44	0.17
	Normalized change scores	0.05	0.04	1.36	0.18	0.30
Scoring explanation plus incentive > Scoring explanation	Posttest items	-0.14	1.53	-0.09	0.93	0.03
	Transfer items	-0.16	0.24	0.67	0.12	0.16
	Game context items	-0.55	0.76	-0.72	0.47	-0.18
	Normalized change scores	0.03	0.04	0.59	0.55	0.17
Scoring explanation plus incentive > Minimal scoring	Posttest items	0.42	1.36	0.26	0.80	0.06
	Transfer items	-0.33	0.21	1.57	0.12	0.36
	Game context items	0.04	0.68	-0.02	0.99	0.00
	Normalized change scores	0.08	0.04	2.05	0.04	0.53

Hypothesis 1. The first hypothesis stated that students in the scoring explanation group would outperform the students in the minimal scoring information groups on the math achievement measures. The results from the comparisons of the two groups on the math achievement measures were not significant.

Hypothesis 2. The second hypothesis stated that students who were in the scoring explanation with incentive would outperform students in the scoring explanation group on the math achievement measures. The results from the planned comparisons were not significant for any of the math achievement measures.

Hypothesis 3. The third hypothesis stated that students in the scoring explanation with incentive would do better than students in the minimal scoring information group on the math achievement measures. Only the planned comparison result for the normalized change scores was significant, $t(109) = 2.06$, $p = .04$, with students in the explanation of scoring rules plus incentive to seek feedback having the higher normalized change scores.

Examining effect of treatment conditions: Adjusting for pretest scores. A linear regression framework was used to compare the differences between providing both the explanation of scoring rules as well as an incentive versus providing minimal to no scoring information on posttest scores after adjusting for pretest scores. For the posttest scores, the model that included both pretest scores and the treatment condition was significant, $F(2, 79) = 173.56$, $p < .001$, explaining 81.0% of the variance. The coefficient ($\beta = 1.24$) for treatment condition was also significant, $t(79) = 2.11$, $p = .04$. However, the treatment condition only explains 1% of the variance. For the transfer item scores, the model that included both pretest scores and the treatment condition was significant, $F(2, 79) = 15.88$, $p < .001$, explaining 28.7% of the variance. However, the coefficient for the treatment condition ($\beta = -.26$) was not significant, $t(79) = 1.46$, $p = .15$.

Interaction between pretest scores and treatment variation. It was also expected that there would be an interaction between prior knowledge (as measured by the pretest scores) and the treatment conditions on the math achievement measures. Three ANCOVAs were computed to test the significance of the interaction between each of the treatment conditions and pretest scores on the posttest items, game context items, and transfer items. The ANCOVA for the interaction between treatment conditions and pretest scores on the game context item scores was significant, $F(2, 106) = 3.27$ $p = .04$, partial eta$^2 = 0.06$. The lines in Figure 7 indicate that for the students with low prior knowledge, the game context item scores were highest for the students in the scoring explanation with incentive. In contrast, for students with high prior knowledge, the game context item scores were lowest for the students in the scoring explanation with incentive.

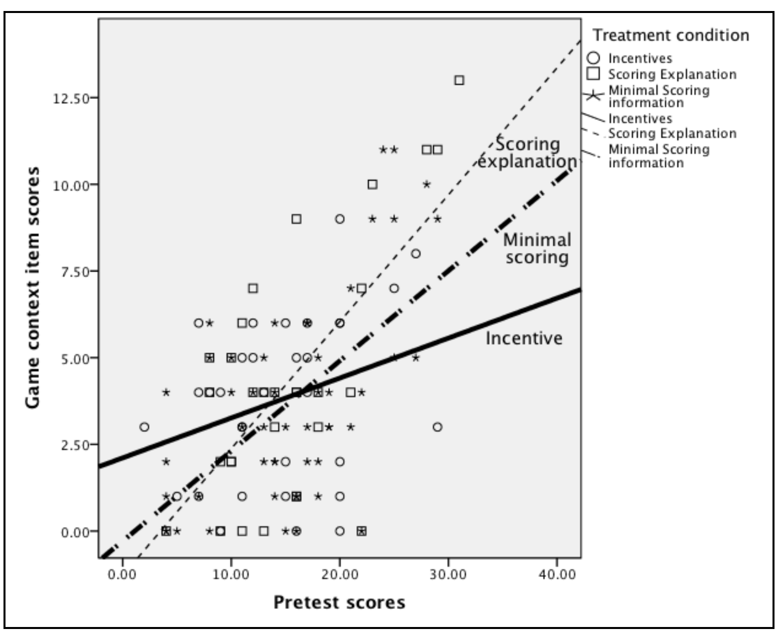


Figure 7. Graph of the interaction effect between pretest scores and treatment variation on the game context item scores.

Results from the Analyses of the Effect of the *Save Patch* Treatments on Game Performance

The next set of results focus on the analyses of the effect of the *Save Patch* treatments on the game performance measures (Research Questions 2b and 3b). First, descriptive information is given, which also includes the rationale for dropping several students' data for one variable. Then, a correlation matrix is presented to provide information about the correlation among the game performance measures and the math achievement measures. Finally, results from inferential analyses of the game performance measures are discussed.

Descriptive information about the game performance measures. The descriptive statistics of the game performance measures are reported in Table 20. For the variable "maximum level reached after 20 minutes" data from eight students were dropped because there was a technical error in the game. The game recorded a maximum level that did not reflect their actual performance. Because their data could not be otherwise verified, the data were dropped. Three of these students were in the scoring explanation with incentive group, three of the students were in the scoring explanation groups and two of the students were in the minimal scoring information group.

Table 20

Descriptive Statistics for the Game Performance Measures Across Students

Game performance measures	Min.	Max.	*M*	*SE*	*SD*	Skewness
Number of coils with unlike denominators added together	0	35	10.82	0.76	8.00	0.97
Number of wrong sized unit coils used	9	71	32.99	1.28	13.57	0.70
Maximum level reached after 20 min.	5	38	15.88	0.49	5.02	1.34
Total number of resets	2	42	10.13	0.56	5.96	2.13
Total number of failed attempts	1	25	6.79	0.38	4.02	1.51

Descriptive statistics were also computed separately for each treatment condition and are reported in Table 21.

Table 21

Descriptives: Game Performance Measures (By Condition)

Game performance measures	Min.	Max.	*M*	*SE*	*SD*	Skewness
Number of coils with unlike denominators added together						
Scoring explanation with incentive	0	20	6.69	0.87	4.92	1.18
Scoring explanation group	1	26	9.63	1.13	6.17	0.77
Minimal scoring information group	0	35	14.18	1.29	9.15	0.49
Number of coils with wrong sized units used						
Scoring explanation with incentive	9	61	30.63	2.08	11.75	0.78
Scoring explanation group	17	71	34.23	2.75	15.04	0.91
Minimal scoring information group	9	64	33.76	1.95	13.82	0.44
Maximum level reached						
Scoring explanation with incentive	10	27	16.03	0.59	3.17	1.29
Scoring explanation group	6	25	15.81	0.98	5.09	0.09
Minimal scoring information group	5	38	15.83	0.85	5.92	1.73
Number of resets						
Scoring explanation with incentive	2	20	7.66	0.64	3.62	1.47
Scoring explanation group	3	16	8.80	0.64	3.49	0.20
Minimal scoring information group	3	42	12.52	1.04	7.34	1.73
Total number of failed attempts						
Scoring explanation with incentive	2	14	5.34	0.45	2.56	1.46
Scoring explanation group	2	11	5.43	0.53	2.88	0.66
Minimal scoring information group	1	25	8.52	0.66	4.70	1.17

The results indicate that students in the scoring explanation with incentive group did better in the game, especially with respect to the number of coils with unlike denominators that were added together. Students in the minimal scoring information group appear to have the weakest game performance, significantly adding more coils with unlike denominators together and resetting the level more often when compared to the other two groups.

Correlations among game performance measures and math achievement measures. Past studies using *Save Patch* have demonstrated that pretest scores explain most of the variance on the game performance measures (Delacruz et al., 2010). Therefore, partial correlations were computed, controlling for pretest (see Appendix N). The results indicate that after controlling for pretest scores, there were no correlations among the game performance measures and math achievement measures.

Effect of the *Save Patch* treatment conditions on game performance. The next set of reported results are from the inferential analyses of the following two research questions:

> Research Question 2b: What is the effect of providing an explanation about the scoring rules on game performance?
>
> Research Question 3b: Does providing an incentive to seek additional feedback affect game performance?

First, ANOVAs were computed to determine if there was an effect of treatment condition on the game performance measures. The three variations of the treatment condition were the three levels of the individual variable. Each measure of game performance was used as separate dependent variables. The results are reported in Table 22. The ANOVAs were significant for most of the game performance measures, with the exception of the maximum level reached in the game.

Table 22

Results from the ANOVAs for the Game Performance Measures

Game performance measures[a]	Source of variance	*SS*	*MS*	*F*
Added coils with different denominator	Regression	1153.21	576.60	10.56
	Residual	5951.22	54.60	
	Total	7104.43		
Used wrong sized coils	Regression	255.04	127.50	.69
	Residual	20199.99	185.32	
	Total	20454.99		
Maximum level reached (after 20 minutes)	Regression	.91	.46	.02
	Residual	2599.71	25.74	
	Total	2600.62		
Total number of resets	Regression	534.49	267.25	8.56
	Residual	3402.50	31.22	
	Total	3936.99		
Total number of failed attempts	Regression	1153.21	576.60	10.56
	Residual	5951.22	54.60	
	Total	7104.43		

[a]*df* = 2 (Between groups), 109 (Within groups), 111 (Total).

Differential effects between conditions on game performance measures. Next, the results from the orthogonal planned comparisons to confirm the three hypotheses pertaining to the game performance measures are presented (see Table 23). The hypotheses are given prior to giving the results of the analyses and each of the hypotheses are discussed individually.

Table 23

Results From Planned Comparisons of Game Performance Measures

Hypothesis tested	Game performance measure	*MD*	*SE*	*t*	*p*	Cohen's *d*
Scoring explanation < Minimal scoring information	Adding coils with unlike denominators	4.34	1.71	2.54	0.01	0.61
	Coils with wrong size units used	0.52	3.17	0.16	0.87	0.03
	Maximum level reached	-0.16	1.21	-0.14	0.89	0.00
	Number of resets	3.67	1.25	2.93	0.00	0.67
	Number of failed attempts	2.88	0.80	3.58	0.00	0.88
Scoring explanation with incentive < Scoring explanation	Adding coils with unlike denominators	2.95	1.42	2.07	0.04	0.54
	Coils with wrong size units used	3.61	3.47	1.04	0.30	0.27
	Maximum level reached	0.22	1.35	-.16	0.87	0.00
	Number of resets	1.14	0.90	1.27	0.21	0.33
	Number of failed attempts	0.09	0.69	0.13	0.90	0.03
Scoring explanation with incentive < Minimal scoring	Adding coils with unlike denominators	7.29	1.55	4.70	0.00	1.09
	Coils with wrong size units used	3.09	3.10	0.99	0.32	0.25
	Maximum level reached	0.57	1.19	0.05	0.96	0.00
	Number of resets	4.81	1.26	3.83	0.00	0.88
	Number of failed attempts	2.97	0.76	3.92	0.00	0.95

Hypothesis 1. The first hypothesis that stated that students in the scoring explanation group would outperform the students in the minimal scoring information group on the game performance measures was partially supported. Students that were in the scoring explanation group added a significantly lower number of coils with unlike denominators, and had fewer resets and failed attempts. The other comparisons were not significant.

Hypothesis 2. The second hypothesis stated that students in the scoring explanation with incentive group would outperform the students in the scoring explanation group on the game performance measures. Only the comparison of the two groups on the number of coils with unlike denominators that were added together produced a significant result.

Hypothesis 3. The last hypothesis that stated students in the scoring explanation with incentive group would outperform the students in the minimal scoring information group on the game performance measures was partially supported by the data. The students in the scoring explanation with incentive group added fewer coils with unlike denominators, and had fewer resets and failed attempts.

Adjusting for prior knowledge. To determine if the hypotheses tested were still valid after adjusting for pretest scores, ANCOVAs were computed using pretest scores as the covariate. The data met both assumptions of ANCOVA. There were no significant differences between the treatment conditions on the pretest scores, which indicated that the data met the assumption of the independence of the covariate and the treatment effect. Second, there was a relationship between pretest scores and the game performance measures for all of the groups, which means that the data met the assumption of the homogeneity of regression slopes.

Table 24

Results from the ANCOVAs for the Game Performance Measures Controlling for Pretest

Game performance measure	*df*	*SS*	*F*	*p*	Partial eta^2
Adding coils with unlike denominators	2.00	529.34	9.47	0.00	0.16
Coils with wrong size units used	2.00	297.55	1.62	0.20	0.03
Maximum level reached	2.00	1.61	0.07	0.94	0.00
Number of resets	2.00	250.77	7.50	0.00	0.13
Number of failed attempts	2.00	123.21	8.73	0.00	0.15

The results from the ANCOVA indicate that after controlling for pretest, there was an effect of treatment condition on the number of coils with unlike denominators that were added together, the number of resets, and the number of failed attempts. No other result was significant.

The planned comparisons were recomputed after controlling for pretest scores. Both the students in the scoring explanation group and the students given the incentive still outperformed the minimal scoring information group for the following measures: (a) total number of coils with unlike denominators added together, (b) number of resets, and (c) number of failed attempts.

Analyses of the Effect of the *Save Patch* Treatments on Voluntarily Seeking Feedback

The next set of results focus on the analyses of the effect of the *Save Patch* treatments on the frequency of voluntarily seeking feedback (Research Questions 4a and 4b). First, descriptive information is given, both across all students and for each individual condition. Then, results from the inferential analyses of voluntarily seeking feedback in the entire game overall are provided.

Dependent measure: Frequency of the voluntary access of feedback in the entire game. First, descriptive statistics of the average number of times students were given the opportunity to seek feedback are reported in Table 25.

Table 25

Number of Opportunities to Seek Feedback (Overall Sample)

Opportunities to seek feedback	Min.	Max.	*M*	*SE*	*SD*
Total across events	7.00	61.00	26.32	3.23	14.06
Event: Adding coils with unlike denominators	1.00	9.00	2.33	0.37	2.23
Event: Used coil with the wrong sized unit	1.00	54.00	14.27	0.93	9.75
Event: Failed attempt	1.00	73.00	9.11	1.33	11.42

Tables 26 shows descriptive statistics of the proportion of times a student voluntarily accessed the feedback in total overall, as well as after three events for the entire sample: (a) added coils with different denominators, (b) used a coil with the wrong size unit for the grid, and (c) failed to save Patch.

Table 26

Proportion of Times Feedback Was Accessed (Overall Sample)

Access of feedback	Min.	Max.	*M*	*SE*	*SD*
Overall total proportion of feedback that was accessed	0.00	0.32	0.08	0.01	0.06
Event: Adding coils with unlike denominators	0.00	0.50	0.02	0.01	0.09
Event: Used coil with the wrong sized unit	0.00	0.50	0.21	0.01	0.16
Event: Failed attempt	0.00	0.46	0.07	0.01	0.13

The data in Table 27 are the descriptive statistics for the frequency of voluntarily accessing the feedback separately for each individual condition.

Table 27

Proportion of Times Feedback Was Accessed (By Condition)

Access of feedback	Min.	Max.	*M*	*SE*	*SD*
Total proportion of times feedback was accessed across all three events					
Scoring explanation with incentive group	0.00	0.32	0.04	0.01	0.08
Scoring explanation group	0.00	0.20	0.11	0.01	0.05
Minimal scoring information group	0.00	0.20	0.08	0.01	0.05
Event: Adding coils with unlike denominators					
Scoring explanation with incentive group	0.00	0.00	0.00	0.00	0.00
Scoring explanation group	0.00	0.00	0.00	0.00	0.00
Minimal scoring information group	0.00	0.50	0.04	0.02	0.13
Event: Used coil with the wrong sized unit					
Scoring explanation with incentive group	0.00	0.22	0.07	0.02	0.09
Scoring explanation group	0.00	0.50	0.33	0.02	0.11
Minimal scoring information group	0.00	0.45	0.22	0.02	0.14
Event: Failed attempt					
Scoring explanation with incentive group	0.00	0.46	0.07	0.02	0.14
Scoring explanation group	0.00	0.33	0.08	0.02	0.13
Minimal scoring information group	0.00	0.43	0.06	0.02	0.12

The data indicate that students in the scoring explanation group tended to access feedback more often. Students in the scoring explanation with incentive group accessed feedback the least. Notably, only students in the minimal scoring information group accessed feedback after they added coils with unlike denominators.

The final set of results is from the inferential analyses that address the fourth research question examining conditions that effect voluntary access to feedback:

> Research Question 4a: Does lack of information affect the frequency with which students voluntarily access feedback?
>
> Research Question 4b: Does providing an incentive to seek additional feedback effect the frequency with which students voluntarily access feedback?

The data were analyzed to test three hypotheses and are reported next. The hypotheses will be presented prior to giving the results of the analyses and each of the hypotheses are discussed individually. Results from the planned comparisons are reported in Table 28.

Table 28

Accessing Feedback: Hypotheses Testing

Measure	*MD*	*SE*	*t*	*p*	Cohen's *d*
Total proportion of times feedback was accessed					
Scoring explanation < Minimal scoring information	0.03	0.01	1.92	0.06	0.56
Scoring explanation with incentive > Scoring explanation	-0.07	0.02	-4.36	0.00	0.57
Scoring explanation with incentive > Minimal scoring information	-0.04	0.01	-2.93	0.00	1.06
Using a coil with the wrong sized unit for the grid					
Scoring explanation < Minimal scoring information	0.11	0.03	3.86	0.00	1.06
Scoring explanation with incentive > Scoring explanation	-0.26	0.03	-10.02	0.00	2.64
Scoring explanation with incentive > Minimal scoring information	-0.15	0.03	-5.79	0.00	1.33

Hypothesis 1. The first hypothesis stated that students in the minimal scoring information group would access the feedback more often than the students in the scoring explanation group. Only students in the minimal scoring information group accessed feedback when coils with unlike denominators were added together. However, compared to students in the minimal scoring information group, students in the scoring explanation condition accessed the feedback more frequently overall, and after the coil with the wrong sized unit for the grid were used.

Hypothesis 2. The second hypothesis stated that students in the scoring explanation with incentive would access the feedback more frequently than students in the scoring explanation group. The students in the scoring explanation with incentive accessed the feedback less frequently overall and also after the coils with the wrong sized unit for the grid were used. The comparison for the number of failed attempts was not significant.

Hypothesis 3. The final hypothesis stated that students who were in the scoring explanation with incentive group would access the feedback more often than students in the minimal scoring information group. Compared to students who received the incentive, students who were given minimal scoring rules information accessed the frequency feedback more often overall and when the coil with the wrong-sized unit was used.

Exploratory Analyses

Additional analyses were computed to examine if there were treatment differences on the posttest and change scores for the following subsamples of students:

- Grade in school (fourth, fifth, and sixth graders)
- Math grade on report card (A, B, C, D, or F)
- Gender
- Self-efficacy (Low or high)
- Math self-concept (Low or high)
- Game experience (Low or high)
- Preferences for cooperative learning (Low or high)
- Preferences for competitive learning (Low or high)

In order to determine if there were treatment differences for these subsamples of students, independent samples *t* tests were computed for each level of the subsample of students (e.g., students who reported low preferences for cooperative learning or students who reported low math self-efficacy).

Grade in school. There were no significant between-condition differences for any of the students with respect to their grade in school.

Math grade on report card (last year). For students who reported that they received a "D" in math on their report card last year (n = 7), those who received both the explanation of the scoring rules and an added incentive to seek additional feedback had significantly higher posttest scores (M = 18.50, SD = 2.12) than students in the other conditions (M = 7.67, SD = 4.16), $t(3) = 3.28$, $p < .05$, Cohen's d = 3.87.

Males. Male students (n = 25) who were given both the explanation of the scoring rules and an added incentive to seek additional feedback had significantly higher change scores (M = 1.14, SD = 2.54), compared to male students who were given only the explanation of the scoring rules (M = -1.91, SD = 3.53), $t(23) = 2.52$, p = .02, Cohen's d = .93.

Low game experience. For students with low game experience (n = 45), those who received both the explanation of the scoring rules and an added incentive to seek additional feedback had significantly higher change scores (M = 1.88, SD = 2.63), compared to similar students in the other conditions (*M* = -.71, *SD* = 1.86), *t*(31) = 3.27, *p*< .001, Cohen's *d* = 1.17.

Low self-efficacy: Posttest scores. For students who reported low self-efficacy (*n* = 44), those who were given both the explanation of the scoring rules and an added incentive to seek additional feedback were found to have higher posttest scores (*M* = 14.94, *SD* = 5.84), than students who were given minimal scoring rules information (*M* = 11.13, *SD* = 4.97), *t*(30) = 1.97, *p* = .06, Cohen's *d* = .88. Those who were given the explanation of the scoring rules also had higher posttest scores (*M* = 16.00, *SD* = 6.58) than students who were given minimal scoring rules information (*M* = 11.13, *SD* = 4.97), *t*(25) = 2.19, *p* = .04, Cohen's *d* = .88.

Low self-efficacy: Change scores. For students who reported low self-efficacy (*n* = 44), those who were given both the explanation of the scoring rules and an added incentive to seek additional feedback had significantly higher change scores (*M* = 1.94, *SD* = 2.68), compared to students in the other conditions (*M* = .26, *SD* = 2.62), *t*(42) = 2.05, *p*< .05, Cohen's *d* = .65.

Low math self-concept. For students who reported low math self-concept (*n* = 41), those who were given both the explanation of the scoring rules and an added incentive to seek additional feedback had significantly higher change scores (*M* = 2.40, *SD* = 2.44), compared to students in the other conditions (*M* = .23, *SD* = 2.55), *t*(39) = 2.66, *p* = 0.01, Cohen's *d* = .89.

Self-reported low preference for cooperative learning. When considering students who reported a low preference for cooperative learning (*n* = 44), those who were given both the explanation of the scoring rules and an added incentive to seek additional feedback had significantly higher change scores (*M* = 2.77, *SD* = 2.49), compared to students in the other conditions (*M* = .23, *SD* = 2.77), *t*(42) = 2.87, *p*< .01, Cohen's *d* = .97.

Summary. These analyses indicate that providing the combined treatment of the explanation of the scoring scheme and the incentive to seek feedback is beneficial for students who reported receiving a "D" in math the previous year, for male students, for those with a low preference for cooperative learning, and for those who report low levels of: (a) game experience; (b) self-efficacy; and (c) math self-concept.

Feedback in Level 2-2

To examine more closely the differential effects of the treatment conditions on the access of feedback after the wrong-sized unit was used[8], three measures were obtained from analysis of the data in Level 2-2:(a) the number of events that preceded the access of feedback, (b) the amount of time that was spent on the feedback, and (c) the number of events between accessing the feedback and completing the level.

An ANOVA was computed to test the effect of the three treatment conditions on each of the three measures. There was no effect of treatment condition on the number of events that preceded first accessing the feedback in Level 2-2, $F(2, 32) = .70$, $p = .50$. There was an effect of treatment condition on the number of events between first accessing the feedback and completing the level, $F(2,32) = 4.27$, $p = .03$, partial eta^2 = .21. Post-hoc results using the Games-Howell[9] test indicated that there were a fewer number of events between first accessing feedback and completing the level for the scoring explanation with incentive group when compared to the scoring explanation group, with results approaching statistical significance (*MD* = 97.92, *SE* = 38.14), $p =.06$.

Accessing the General Help Menu

An additional analysis of the effect of the treatment conditions on accessing feedback, for the topic[10] "Coil size" was performed. The result from the Chi-Square test was significant, $\chi^2(2) = 6.86$, $p = .03$, indicating that students in the scoring explanation with scoring explanation with incentive group were more likely to access the topic "Coil Size" and students in the minimal scoring group were less likely to access it.

Table 29

Access of General Help Menu: "Coil Size"

Condition	Did not access "Coil Size" topic	Accessed "Coil Size" topic
Scoring explanation with incentive	9	15
Scoring explanation	10	10
Minimal scoring information	23	9
Total	42	34

[8] Analysis of accessing feedback after the wrong-sized unit was used was chosen because it was only after this event in Level 2-2 that all of the treatment variations accessed the feedback.
[9] Levene's test of equality of error variances was significant.
[10] All of the other topics in the general help menu were either accessed only once or twice, or not accessed at all.

Self-Reported Motivation to Play the Game

This analysis was performed to determine if adding the scoring rules information and providing the incentive to voluntarily access feedback help resulted in more positive or negative perceptions of the game. Results from the ANOVAs show that there is no significant effect of treatment condition on either the positive comments, $F(2, 92) = .16$, $p = .85$, or the negative comments, $F(2, 97) = .04$, $p = .95$.

Discussion

This study investigated, experimentally, the impact of game design features in a short mathematics game intended for upper elementary and middle school students. Specifically, the study examined the effects of combining different levels of feedback, including explicit scoring rules and incentives to seek greater feedback on math achievement measures, game play, and help-seeking behaviors.

Limitations

Before discussing the interpretations and implications drawn from the study's results, let us consider four clusters of limitations of the study: (1) study design, (2) treatment conditions, (3) game design, and (4) number of statistical tests.

Study Design

- The study focused on a single game, albeit one with experimental variations. The game is only one of a large number that might have been used in this study and findings would have to be replicated for other similar games.
- The procedures of the study were constrained by available time in the after-school settings. The activities required in the study were a pretest, game play, and a posttest, totaling about 90 minutes. Experimental students, therefore, may not have given the posttest their full effort due to fatigue or repetition of items, and results may have underestimated how much was learned from playing the game.

Treatments

- The participants in the experimental conditions played the game for about 40 minutes between taking the pretest and the posttest. Also, treatment variations were triggered only when a mistake was made. So the total opportunities to experience the assigned treatment were limited.
- The incentive option used the "recapturing" of lost points, after an error was made. In *Save Patch*, points were not intrinsic to the game process, but served only as an external indicator of competence to the students. For example, accumulated points did not give players more resources to use in the game and influence their proficiency.

Game Design

- All experimental participants were exposed to a tutorial on game content and process. The tutorial itself could have obscured treatment differences, because it contained specific, relevant instruction that was taught in the game.

Total Number of Statistical Tests

- The analysis of study results required many statistical tests, involving the main and more specific research questions and multiple dependent measures. Numerous statistical tests raise the general concern of Type I Error.

Interpretation of Findings

Playing *Save Patch* Compared to a Control

The findings from the current study demonstrate that, after adjusting for pretest scores, playing *Save Patch* leads to higher math achievement (i.e., higher posttest and higher normalized change scores) when compared to a control group. However, the differences obtained were small. While the findings from this study are mildly promising, *Save Patch* is not as of yet an effective learning environment. *Save Patch* was, in fact, designed to be a test bed for empirical research on different design variables, rather than a game fully ready for implementation in a classroom. A highly effective game would have presented the likelihood of ceiling effects.

One potential explanation for not finding stronger differences between playing *Save Patch* and the control group was the measure of transfer. *Save Patch* was not intended to teach a procedure but to support conceptual learning, best measured by transfer items. However, analysis of the four transfer items, a part of the posttest, produced a low Cronbach's alpha coefficient, limiting inferences from this measure.

An additional troubling problem is that the correlations between pretest and posttest were high and that the actual score values before and after the game were similar. Although pretest performance always predicts posttest scores, it is clear that the subjects as a group did not have much math expertise (answering approximately 50% correct on the pretest and marginally more on the posttest). Taken together with generally weak findings, these relationships may suggest that the sample in the study was not ideally suited for the game goals.

Differential Effects of Treatment Conditions on Math and Game Performance

The data did not support the hypothesis that just providing the explanation of scoring rules would lead to better performance on the math achievement measures. Although students who were given the full explanation of scoring rules accessed the feedback more often, they did not learn more than the students who got minimal information. This may be because the students in the explanation of scoring rules group were unable to or elected not to process the information closely.

It is puzzling that no differential impact on learning was found for the students who received the explanation of scoring rules only because they accessed the feedback most often and spent the most time on it. The combined explanation of scoring rules and the incentive treatment was observed to increase math learning, resulting in larger changes between the pretest and posttest in comparison to the other treatments. Yet, the students given the incentive (recapturing points by accessing feedback) actually accessed the feedback less and spent less time on it. A potential explanation for this finding may be considered. It is possible that incentives combined with the more extended scoring explanation signaled the importance of the feedback and increased the effectiveness of feedback use because of heightened attention of deeper processing. This explanation is obviously speculative since there were no direct measures of the subjects' attention to the feedback screens or of their depth of processing.

Additionally, a treatment interaction was found on the game context item scores (items like those played in the game). For students with lower pretest scores, providing the incentive plus explanation of rules led to higher scores on the game context items. Students who began with higher pretest scores did better with just the scoring explanation. This finding may have occurred because students with higher prior knowledge were better students and therefore able to make sense of the scoring explanation and use it to repair mistakes.

In contrast, students with lower prior knowledge needed both the additional information provided by the feedback and the extra push to use it when necessary. The explanation that students with lower prior knowledge may need both the additional information and the incentive to seek additional feedback when necessary is supported by the results from the analyses of specific subsamples. Providing both the explanation of the scoring rules and offering an incentive to seek additional feedback is more effective for students who reported receiving a "D" in math on their report card the previous year, as well as students who reported low game experience, low self-efficacy, low math self-concept, and low preferences for cooperative learning environments.

The variables self-efficacy, math self-concept, and preferences for cooperative learning were significantly correlated. This suggests they all may be tapping into a similar construct, which may be low academic intrinsic motivation. Academic intrinsic motivation has been conceptualized as the intrinsic motivation to learn (Brophy, 1983), and has been found to be negatively correlated with both self-perceptions of competence (i.e., self-efficacy) and academic achievement, especially in the area of math (Gottfried, 1985). Similarly, it has been argued that extrinsic rewards, especially those that are not contingent on performance, are more beneficial for students who do not care about academic work or have had a history of failure which characterize students with low academic motivation (Sternberg & Williams, 2002). These findings provide further support that in games for learning, both providing explicit information about the scoring rules and incentivizing the use of feedback are beneficial, particularly for students who typically underperform in school.

Implications of the Study for Game Design and Future Research

This study sought to improve understanding of what attributes make games more effective. If the game were transformed from a test bed to a usable game, its efficacy in larger trials (comparing its use in supervised or unsupervised settings) would address the extent to which the game itself extends time on task and is sufficiently motivating. This raises game design implications for research.

Redesign of the tutorials could be explored by considering different approaches to providing information more "native" to the game. This approach implies that the tutorial not hang outside of the game as "instruction" but be more integrated into early game play. Second, the reading load of the tutorial will need adjustment if the game's target sample is to remain underperforming students. Third, to further support game play for students without the requisite prior knowledge, two additional options might be considered in a revision: (1) revising the game itself so that it begins with lower level concepts to enable students to profit from the game is one option; and (2) providing priming or warm-up experience intended to stimulate students to activate knowledge relevant for game learning and success may be a useful option for future research.

If one goal of instruction is to facilitate understanding of similarity and differences among classes of problems—adoption of a systematic, explicit approach to schema development such as analogical reasoning (Holyoak, Gentner, & Kokinov, 2001) may also be explored. In *Save Patch*, students were given only one type of problem to solve: compute the distance of the jump and use fractional components to solve the problem. Students were not given opportunities to solve different problems that used the same concepts. With the purpose of creating cheats or walkthroughs (i.e., guides for players to support game play), students could be asked to identify the elements of each case and create a relational mapping between them, identifying the commonalities or differences between the game examples

(Catrambone & Holyoak, 1989; Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983; Novick & Holyoak, 1991).

While the comparisons between the experimental and control groups and pre- to posttest changes provide, at best, modest evidence of the effectiveness of the game, findings from the treatment variations may suggest features to explore in the design of learning games, specifically variations in feedback and incentives. Because students infrequently use available help, such as feedback in this study (Aleven et al., 2003; Conati& Zhao, 2004; Nelson, 2007; Van Eck & Dempsey, 2002), there is a need to examine ways that motivate students to seek help. Although this study stimulated help following an error, research could explore alternative placement of help (e.g., before beginning the problem, as a way to consolidate knowledge). Additionally, feedback might be given on a random schedule to correct answers and as needed on errors.

One possible concern in the use of incentives in this study was the use of negative reinforcement; that is, giving back some "lost points" if feedback was sought after an error rather than a more straightforward reward of positive behavior. In contrast to this procedure, positive incentives are consistent with research on the use of rewards for learning following desired behaviors (Holland & Skinner, 1961). A study that provided positive incentives may be more worth exploration.

Another concern is that in game playing, there is an ethos to explore the domain without reading instructions or seeking help. In addition, the "discourse" of gaming often rewards speed, and seeking help or processing feedback runs counter to this discourse by slowing things down. Changing the game so that accrued points could "buy" functional tools for the student should be explored, so that the incentive connected functional rather than symbolic value for the player. Points could be made intrinsic to the game in a number of ways, for instance, by allowing players to acquire useful material goods, gain power that provides more strength, or develop additional skills that facilitate success in the game (Bostan, 2009). This may make the actual incentive more valuable to the player beyond a monitoring of attainment. Experimental variations of intrinsic point value could be explored.

Any future study will need to review carefully test items related to game goals and transfer items. At minimum there is a need for investigation of how the assessment works for students who differ in prior knowledge and other background characteristics. Another potential validity issue would be to test the instructional sensitivity of the measure in experimental but not necessarily game-based instructional settings.

Also, more items and different item types are needed to assess transfer, and to explore the multidimensionality of task requirements. Such an analysis will undoubtedly result in a wider range of item types in a longer game. Clearly if transfer were to remain an important outcome, more items of high technical quality would be needed for its measurement in any future study.

In conclusion, in order to determine whether or not games teach academic content and skills requires the collection of empirical evidence, the use of stronger methodologies in the study designs, and a close examination of the different game design features that may lead to learning. Like formative assessment in non-game settings, more evidence is needed to determine whether and how criteria of performance and scoring rules should be communicated to students to be useful for learning. Given that the incentive to seek additional feedback was most beneficial for students with lower prior knowledge, low self-efficacy, and low math self-concept—additional effort is necessary to examine different approaches in motivating students to use provided feedback through the use of incentives, and investigate how incentives may be made more effective for learning in games.

References

Airasian, P. W., & Jones, A. M. (1993). The teacher as applied measurer: Realities of classroom measurement and assessment. *Applied Measurement in Education, 6*, 241-254.

Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 292-303). Berlin: Springer Verlag.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research, 73*(3), 277-320.

Ayres, P., & van Gog, T. (2009). State of the art research into cognitive load theory. *Computers in Human Behavior, 25*(2), 253-257.

Backlund, P., Engström, H., Johannesson, M., Lebram, M., & Sjödén, B. (2008). Designing for self-efficacy in a game based simulator: An experimental study and its implications for serious games design. *Proceedings of the International Conference Visualisation (Viz08),* 106-113.

Baker, E. L. (1974). Formative evaluation of instruction. In W. J. Popham (Ed.), *Evaluation in education: Current applications* (pp. 531-585). Berkeley, CA: McCutchan.

Baker, E. L., Abedi, J., Linn, R. L., & Niemi, D. (1995). Dimensionality and generalizability of domain-independent performance assessments. *Journal of Educational Research, 89*, 197–205.

Baker, E. L., & Delacruz, G. C. (2007). A framework for the assessment of learning games. In H. F. O'Neil & R. S. Perez (Eds.), *Computer games and team and individual learning* (pp. 21-37). Oxford, UK: Elsevier.

Bangert-Drowns, R. L., Kulik, C.-L., C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213-238.

Barab, S. A., & Dede, C. (2007). Games and immersive participatory simulations for science education: An emerging type of curricula. *Journal of Science Education and Technology, 16*(1), 1-3.

Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S., & Warrant, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology, 18,* 305-320.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, & Practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-144.

Black, P., & Wiliam, D. (1998c). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*, 5-31.

Björk, S., & Holopainen, J. (2005). Games and design patterns. In K. Salen & E. Zimmerman (Eds.), *The game design reader: A rules of play anthology* (pp. 410-437). Cambridge, MA: MIT Press.

Bostan, B. (2009). Player motivations: A psychological perspective. *Computers in Entertainment, 7*(2), 1-32.

Brophy, J. E. (1983). Conceptualizing student motivation. *Educational Psychologist, 18,* 200-215.

Brown, G.T.L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system, *Assessing Writing, 9*, 105–121.

Brown, G., & Quinn, R. J. (2006). Algebra students' difficulty with fractions. *Australian Mathematics Teacher, 62*(4), 28–40.

Buckingham, D., & Scanlon, M. (2003). Interactivity and pedagogy in 'Edu-tainment' software. *Information Technology, Education and Society, 4*(2), 107-126.

Cameron, J. (2001). Negative effects of reward on intrinsic motivation—A limited phenomenon: Comment on Deci, Koestner, and Ryan (2001). *Review of Educational Research, 71*(1), 29-42. Retrieved from http://www.unco.edu/cebs/psychology/kevinpugh/motivation_project/resources/cameron01.pdf

Cameron, B., & Dwyer, F. (2005). The effect of online gaming, cognition and feedback type in facilitating delayed achievement of different learning objectives. *Journal of Interactive Learning Research, 16*(3), 243-258. Retrieved from http://www.thefreelibrary.com/_/print/PrintArticle.aspx?id=135121407

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research, 64*(3), 363-423.

Catrambone, R., & Holyoak, K. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 1147-1176.

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). Aligning game mechanics and learning goals. 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 American Educational Research Association conference, Denver, CO.

Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessmen t*(pp.3-17). New York, NY: Routledge.

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research And Development, 42*(2), 21-29.

Clark, D. B., Nelson, B., Sengupta, P., & D'Angelo, C. M. (2009). *Rethinking science learning through digital games and simulations: Genres, examples, and evidence.* Invited Topic Paper in the Proceedings of The National Academies Board on Science Education Workshop on Learning Science: Computer Games, Simulations, and Education. Washington, DC. Retrieved from http://www7.nationalacademies.org/bose/Clark_Gaming_CommissionedPaper.pdf

Clark, R. C., & Mayer, R. E. (2007). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning, 2nd edition.* Hoboken, NJ: John Wiley & Sons.

Conati, C., & Zhao, X. (2004). Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game. *Proceedings of the 9th International Conference on Intelligent User Interfaces*, 6-13.

Council of Chief State School Officers. (2010). *Model core teaching standards: A resource for state dialogue*. Retrieved from the CCSSO website http://www.ccsso.org/Documents/2010/Model_Core_Teaching_Standards_DRAFT_FOR_PUBLIC_COMMENT_2010.pdf

Davidson, L., & Scripp, L. (1990). Tracing reflective thinking in the performance ensemble. *The Quarterly, 1*(1&2), 49-62.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125*, 627–668.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research, 71*(1), 1-27.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227-268. Retrieved from http://updatenet.ch/images/7/75/Selbstbestimmung_Deci_Ryan_2000.pdf

De Freitas, S. L. (2006). Using games and simulations for supporting learning. *Learning, Media, and Technology, 31*(4), 343-358.

Delacruz, G. C., Chung, G. W. K., & Baker, E. L., (2010). Validity evidence for games as assessment environments. In K. Gomez, L. Lyons, & J. Radinsky (Eds.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 2, Short Papers, Symposia, and Selected Abstracts.* Chicago, IL: International Society of the Learning Sciences.

Dempsey, J. V., Rasmussen, K., & Lucassen, B. (1994, February). Instructional gaming: Implications for instructional technology. Paper presented at the annual meeting of the Association for Educational Communications and Technology, Nashville, TN.

Dieterle, E. (2009). Neomillennial learning styles and River City. *Children, Youth and Environments, 19*(1), 245–278.

Driscoll, M. (1982). *Research within reach: Secondary school mathematics. A research guided response to the concerns of educators.* St. Louis, MO: CEMREL.

Egenfeldt-Nielsen, S. (2005). *Beyond edutainment: Exploring the educational potential of computer games* (Unpublished doctoral dissertation). IT University of Copenhagen, Denmark.

Fisher, C., Filby, N., Marliave, R., Cahen, L., Dishaw, M., Moore, J., & Berliner, D. (1978). *Teaching behaviors: Academic Learning Time and student achievement* (Final report of Phase III-B, Beginning Teacher Evaluation Study). San Francisco, CA: Far West Laboratory for Educational Research and Development.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27-32.

Fullerton, T. (2008). *Game design workshop: A playcentric approach to creating innovative games* (2nd ed.). New York, NY: Morgan Kaufmann.

Fuson, K. C. (2003). Research on whole number addition and subtraction. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 243-275). Reston, VA: National Council of Teachers of Mathematics.

Gearhart, M., Herman, J. L., Novak, J. R., & Wolf, S. A. (1995). Toward the instructional utility of large-scale writing assessment: Validation of a new narrative rubric. *Assessing Writing, 2*, 207–242.

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

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393-408.

Gick, M. L., & Holyoak, K. (1983). Schema induction and analogical transfer.*Cognitive Psychology, 15,* 1-38.

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology, 6,* 631-645.

Green, R., & Bowser, M. (2006). Observations from the field. *Journal of Library Administration, 45*(1), 185-202.

Greenblat, C. S. (1973). Teaching with simulation games: A review of claims and evidence. *Teaching Sociology, 1*(1), 62-83. Retrieved from JSTOR website http://www.jstor.org/stable/1317334

Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*, 64-74.

Hays, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion* (Technical Report). Retrieved from http://adlcommunity.net/file.php/36/GrooveFiles/Instr_Game_Review_Tr_2005.pdf

Hattie, J., & Temperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Heritage, M., Kim, J., Vendlinksi, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28*(3), 24-31.

Hickey, D. T., Ingram-Goble, A., & Jameson, E. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education Technology, 18*, 187-208.

Holland, J. G., & Skinner, B. F. (1961). *The analysis of behavior: A program for self-instruction*. New York, NY: McGraw-Hill.

Holyoak, K. J., Gentner, D. & Kokinov, B. N. (2001). Introduction: The place of analogy in cognition. In D. Gentner, K. J. Holyoak, & B. N. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 1-19). Cambridge, MA: MIT Press.

Hounsell, D. (1997). Contrasting conceptions of essay writing. In F. Marton, D. Hounsell, & N. J. Entwistle (Eds.), *The experience of learning: Implications for teaching and studying in higher education* (2nd ed., pp. 106–125). Edinburgh: Scottish Academic Press.

Huizinga, J. (1938). Nature and significance of play as a cultural phenomenon. In K. Salen & E. Zimmerman (Eds.): *The game design reader: A rules of play anthology* (pp. 96-119). Cambridge, MA: MIT Press.

Iseli, M. R., Koenig, A. D., Lee, J. J., & Wainess, R. (2010). *Automated assessment of complex task performance in games and simulations* (CRESST Report 775). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Jenkins, H. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century.* Cambridge, MA: The MIT Press.

deJong, T., & van Joolingen, W.R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68, 179-202.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity, and educational consequences. *Educational Researcher Review, 2*, 130-144.

Karabenick, S. A., & Newman, R. S. (Eds.). (2006). *Help seeking in academic settings: Goals, groups and contexts*. Mahwah, NJ: Erlbaum.

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In R. E. Ferdig (Ed.), *Handbook of research on effective electronic gaming* (Vol. 1, pp. 1-32). Hershey, PA: Information Science Reference.

Ke, F., & Grabowski, B. (2007). Gameplaying for math learning: Cooperative or not? *British Journal of Educational Technology, 38*(2), 249-259.

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.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up*. Washington, DC: National Academy Press.

Kirriemuir, J., & McFarlane, A. (2004). Literature review in games and learning [Electronic Version]. *Future Lab Serie*s, 1-35. Retrieved from http://www.futurelab.org.uk/resources/documents/lit_reviews/Games_Review.pdf

Klopfer, E., & Squire, K. (2008). Environmental detectives—The development of an augmented reality platform for environmental simulations. *Educational Technology Research and Development, 56*(2), 203-228.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Koretz, D., Klein, S., McCaffrey, D., & Stecher, B. (1993). *Interim report: The reliability of Vermont portfolio scores in the 1992-93 school year* (CSE Technical Report 370). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Koster, R. (2005). *A theory of fun.* Scottsdale, AZ: Paraglyph Press.

Laffey, J. M., Espinosa, L., Moore, J., & Lodree, A. (2003). Supporting learning and behavior of at-risk young children: Computers in urban education. *Journal of Research on Technology in Education, 35*(4), 423-440. Retrieved from https://www.msu.edu/~shubelro/researcharticle.pdf

Lamon, S. J. (1999). *Teaching fractions and ratios for understanding: Essential knowledge and instructional strategies for teachers.* Mahwah, NJ: Erlbaum.

Lee, J., Luchini, K., Michael, B., Norris, C., & Soloway, E. (2004). *More than just fun and games: Assessing the value of educational video games in the classroom*. Paper presented at the CHI '04 Extended Abstracts on Human Factors in Computing Systems, Vienna, Austria.

Lumsdaine, A. (1963). Instruments and media of instruction. In N. Gage (Ed.), *Handbook of research on teaching*. Chicago, IL: Rand McNally.

Malone, T. W. (1980). What makes things fun to learn? Heuristics for designing computer games. *Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on small systems* (pp. 162-169), New York, NY: ACM.

Malone, T. W. (1981). What makes computer games fun? *Byte, 6*(12), 258-277.

Markle, S. M. (1967). Empirical testing of programs. In P. C. Lange (Ed.), *Programmed instruction* (The sixty-sixth yearbook of the National Society for the Study of Education, Part II, pp. 104-138). Chicago, IL: National Society for the Study of Education.

Marsh, H. W., Hau, K. T., Artlet, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing, 6,* 311-360.

Marx, J., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, *75*, 87-91.

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

McCreery, M.P., Krach, S.K., & Schrader, P.G. (2010). *Experts and novices in virtual space: An investigation of issues in education*. Manuscript submitted for publication.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1,* 30-46.

Mielke, K. (1968). Questioning the questions of ETV research. *Educational Broadcasting Review*, *2*, 6-15.

Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. *Cognitive Science: A Multidisciplinary Journal, 23*(3), 305-336.

Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology, 33*(2), 270-298.

Mor, Y., Winters, N., Cerulli, M., & Björk, S. (2006). *Literature review on the use of games in mathematical learning, Part I: Design. Report of the Learning Patterns for the Design and Deployment of Mathematical Games project.* Retrieved from http://lp.noe-kaleidoscope.org/outcomes/litrev/LP-LitReview-v2-desdep-20060314.pdf

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science, 32*(1/2), 99-113.

Moreno, R. (2009). Cognitive load theory: More food for thought. *Instructional Science, 38*(2), 135-141.

Mullen, Y. K. (2003). *Student improvement in middle school science* (Unpublished master's thesis). University of Wisconsin.

National Research Council. (2001). *Classroom assessment and the National Science Education Standards. Committee on Classroom Assessment and the National Science Education Standards*. J. M. Atkin, P. Black, & J. Coffey (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2003). *Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment.* Workshop report: Committee on Assessment in Support of Instruction and Learning, Board on Testing and Assessment, Committee on Science Education K-12, Mathematical Sciences Education Board, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

Nelson, B. C. (2007). Exploring the use of individual, reflective guidance in an educational multi-user virtual environment. *Journal of Science Education and Technology, 16*(1), 83-97.

Neulight, N., Kafai, Y. B., Kao, L., Foley, B., & Galas, C. (2007). Children's participation in a virtual epidemic in the science classroom: Making connections to natural infectious diseases. *Journal of Science Education and Technology, 16*(1), 47-58.

Norton, L. S. (1990). Essay writing: What really counts? *Higher Education, 20*(4), 411–442.

Novick, L. R. & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(3), 398-415.

Nyquist, J, B. (2003). *The benefits of reconstruing feedback as a larger system of formative assessment: A meta-analysis* (Unpublished master's thesis).Vanderbilt University, Nashville, TN.

Okan, Z. (2003). Edutainment: Is learning at risk? *British Journal of Educational Technology, 34*(3), 255-264.

O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: evidence from the computer games literature. *Curriculum Journal, 16*(4), 455-474.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education, 52*, 1-12.

Paas, F., van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, *22*, 115-121.

Plaas, J. L., Homer, B. D., Milne, C., Jordan, T., Kalyuga, S., Kim, M., & Lee, H. (2009). Design factors for effective science simulations: Representation of information. *International Journal of Gaming and Computer-Mediated Simulations, 1*, 16-35.

Popham, W. J. (2006). *Defining and enhancing formative assessment.* Washington, DC: Council of Chief State School Officers.

Presnky, M. (2005). Computer games and learning: Digital game-based learning, In J. Raessens & J. Goldstein (Eds.), *A handbook of computer game design studies* (pp. 97-124). Cambridge, MA: MIT Press.

Puustinen, M., Volckaert-Legrier, O., Coquin, D., & Benicot, J. (2009). An analysis of students' spontaneous computer-mediated help-seeking: A step toward the design of ecologically valid supporting tools. *Computers & Education, 53*(4), 1040-1047.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development, 44*(2), 43-58. Retrieved from http://www.jstor.org/stable/30221022

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context.* New York, NY: Oxford University Press.

Rogoff, B. (2003). *The cultural nature of human development.* New York, NY: Oxford University Press.

Rosas, R., Nussbaum, M.,Cumsille, P., Marianov, V., Correa, M., Flores, P., & Salinas, M. (2003). Beyond Nintendo: Design and assessment of educational video games for first and second grade students. *Computers & Education, 40*, 71-94.

Rosenbaum, E., Klopfer, E., & Perry, J. (2007). On location learning: Authentic applied science with networked augmented realities. *Journal of Science Education and Technology, 16*(1), 31-45.

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *Journal of Technology, Learning, and Assessment, 8*(4). Retrieved from http://escholarship.bc.edu/jtla/vol8/4

Ryan, A. M., & Pintrich, P. R. (1997). "Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology, 89,* 329-341.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy, & Practice, 5*(1), 77-84.

Sadler, P. M. & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1-31.

Salden, R., Aleven, V., & Renkl, A. (2007). Can tutored problem solving be improved by learning from examples*? Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 1847.

Salen, K., & Zimmerman, E. (2005). *Rules of play: Game design fundamentals*. Cambridge, MA: The MIT Press.

Saxe, G. B., Gearhart, M., & Seltzer, M. (1999). Relations between classroom practices and student learning in the domain of fractions. *Cognition and Instruction, 17,* 1–24.

Schell, J. (2005). *The art of game design: A book of lenses.* Burlington, MA: Morgan Kaufmann Publishers.

Schwonke, R., Wittwer, J., Aleven, V., Salden, R. J. C. M., Krieg, C., & Renkl, A. (2007, May). *Can tutored problem solving benefit from faded worked-out examples?* Paper presented at The European Cognitive Science Conference, Delphi, Greece.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation* (American Educational Research Association Monograph Series on Curriculum Evaluation, No. 1, pp. 39-83). Chicago, IL: Rand McNally.

Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers and Education. 46*(3), 223-234. Retrieved from http://epistemicgames.org/cv/papers/shaffer_cae_2005.pdf

Shaffer, D.W., & Gee, J. P. (2007). Epistemic games as education for innovation. In J.D.M. Underwood &J. Dockrell (Eds.), *Learning through digital technologies* (pp. 71-82). Leicester, UK: British Journal of Educational Psychology. Retrieved from http://epistemicgames.org/cv/papers/Ep-games-innovation-Shaffer-Gee-BPS-2007.pdf

Shaffer, D.W., Hatfield, D., Svarovsky, G.N., Nash, P., Nulty, A., Bagley, E., & Mislevy, R.J. (2009). Epistemic network analysis: A prototype for 21st century assessment of learning. *International Journal of Learning Media, 1*(2), 33–53.

Shaffer, D. W., Squire, K., Halverson, R., & Gee, J. (2005). Video games and the future of learning. *The Phi Delta Kappan, 87*(2), 104-111. Retrieved from http://www.jstor.org/stable/20441942

Shute, V. J. (1993). A comparison of learning environments: All that glitters. In S. P. Lajoie & S. J. Derry (Eds.), *Computers as cognitive tools* (pp. 47-75). Hillsdale, NJ: Erlbaum.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Spivey, P.M. (1985). *The effects of computer assisted instruction on student achievement in addition and subtraction at first-grade level*. (ERIC Document No. ED 263874). Retrieved from ERIC database

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher, 35*(8), 19-29.

Squire, K., & Jan, M. F. (2007). Mad city mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology, 16*(1), 5-29. Retrieved from website.education.wisc.edu/kdsquire/.../madcity-squire-jan-final.pdf

Sternberg, R., & Williams, W. (2002). *Educational Psychology*. Boston, MA: Allyn and Bacon.

Stiggins, R. J. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*(4), 324-328. Retrieved from http://www.jstor.org/stable/20441998

Supovitz, J. A., MacGowan, A., & Slattery, J. (1997). Assessing agreement: An examination of interrater reliability of portfolio assessment in Rochester, New York. *Educational Assessment, 4*(3), 237-259.

Svarovsky, G. N. (2010). *Unpacking an engineering practicum: Building engineers, one participant structure at a time* (WCER Working Paper No. 2010-5). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Taras, M. (2010). Assessment for learning: Assessing the theory and evidence. *Procedia Social and Behavioral Sciences, 2*, 3015-3022.

Tobias, S., & Fletcher, J.D. (2007). What research has to say about designing computer games for learning. *Educational Technology, 47* (5), 20-29.

Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841–873). New York, NY: Macmillan.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition & Instruction, 10*(1), 1-100.

Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment, 11*(3/4), 283-289.

Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). New York, NY: Routledge.

Wu, H. (2001). How to prepare students for algebra. *American Educator, 25*(2), 10–17.

Van Eck, R., & Dempsey, J. (2002). The effect of competition and contextualized advisement on the transfer of skills in a computer-based instructional simulation game. *Educational Technology, Research, and Development, 50*(3), 23-41.

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

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34*(3), 229-243.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 38*(9), 25-32.

Appendix A:
Sample Assessment Architecture (*Save Patch*)

Cognitive demand (Problem solving)	Domain Representation (Addition of rational numbers)	Task specification (*Save Patch* game design)
Goal(s): Determine goal state	Goal(s): If given a set of fractions, figure out what fractions need to be added together to produce a specific quantity.	Goal(s): Compute the distance between T-blocks. Estimate the number of coils needed. Select the size of coil needed.
Givens: What pieces of information or data have been given?	Givens: Boundaries of a whole unit	Givens: Depicted by intersection of red lines
	Number of intervals whole unit has been broken into	Jump space, coil size
Parameters: What are the constraints or rules of the problem space?	Parameters: The size of fraction is relative to whole unit.	Parameters: In the game, the denominator of the jump space = number of pieces between red lines.
	Like whole number integers, fractions with like denominators can be added together to produce a given quantity.	Total jump space = number of pieces between T-blocks. Quantity of coils needed = size of jump space.
	A quantity can be decomposed into a number of smaller, equal pieces.	Coils can be broken down into smaller, equal-sized pieces.
	Only fractions with common denominators can be added together.	Coils that are the same size can be combined on a trampoline.
	Fractions with uncommon denominators must be converted to the same unit before they can be added together.	Coils with different sizes cannot be added together.
Select the appropriate solution: How can the goal be achieved?	Construct the appropriate solution: If presented with fractions with common denominators, figure out how many of them need to be added together to equal a given quantity.	Construct the appropriate solution: To determine the quantity of coils needed, determine the size of the jump space and decide how many of the coils provided need to be added together so that they equal the jump distance.
	Example: Size of jump space = 6/4 Coils available: Eight ¼ unit coils Solution: Add six ¼ unit coils.	
	If presented with fractions with uncommon denominators, convert fractions so that they have a common denominator.	Change the size of the coils so that they have the same denominator.
	Example: Size of jump space = 3/4 Coils available: One ¼ unit coil and one ½. Solution: Convert the ½ coil into two ¼ unit coils and add three ¼ unit coils together.	

Appendix B:
Rational Number Knowledge Specifications

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 (the 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.

4.2.0. If the numerator is smaller than the denominator, the fraction represents a number less than one whole unit.

4.3.0. If the numerator is equal to the denominator, the fraction represents one whole unit.

4.4.0. 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.

6.0.0. Does the student understand the meaning of negative numbers?

6.1.0. Negative numbers are those numbers that are opposites (direction, side, etc.) of the positive numbers.

6.2.0. Zero separates the positive from the negative numbers.

6.2.1. Zero is neither positive nor negative.

6.2.2. Zero has no opposite.

6.3.0. The placement of zero depends on context or is relative to some benchmark (e.g. 0 degrees Celsius is the freezing point of pure water, etc.).

6.3.1. In an absolute system, there are no negative numbers.

6.3.2. If zero is placed so that numbers less than zero are possible, these numbers are called negative numbers.

6.4.0. The magnitude (absolute value) of negative and positive numbers increases as their distance from zero increases.

6.5.0. The value of a negative number decreases as its distance from zero becomes larger.

7.0.0. Does the student understand the meaning of multiplication?

7.1.0. Multiplication represents a number of groups of identical quantities.

7.2.0. In two factor multiplication, one factor (the multiplier) shows the number of groups; the other factor (the multiplicand) shows the identical quantity in each group.

7.3.0. Multiplication is commutative so any number can represent the number of groups and other multiplicand(s) represent the number of identical things in each group.

7.4.0. Multiplying a quantity by any form of one may change the appearance of a quantity but will not change the value of that quantity (the Multiplicative Identity).

7.5.0. Multiplication of a negative number by a positive integer can be thought of as a number of groups of that negative number.

7.6.0. Multiplication of a negative number by a positive integer can be thought of as adding in the opposite direction (or the opposite of) adding in the positive direction.

7.7.0. Multiplication of a number by negative one (-1) produces the opposite of that number.

7.7.1. Any number can be factored into the multiplication of that number's opposite and negative one (-1). For example (-3) = (-1)(3) and (3) = (-1)(-3).

8.0.0. Subtracting one number (x) from another number (y) is equivalent to adding the opposite (e.g. y - x = y + -x).

8.1.0. If one adds zero, in the form of a subtrahend and its opposite, to a number, one can change subtraction into addition of the opposite.

8.2.0. Subtraction and addition can be thought of as opposite operations.

8.3.0. Subtraction can be thought of as "taking away" one quantity from another.

8.4.0. Subtraction can be thought of as the "difference" between two numbers.

9.0.0. Dividing one number (x) by another non-zero number (y) is equivalent to multiplying the first number by the reciprocal of the second number (e.g. $x \div y = x \bullet 1/y$).

9.1.0. The product of multiplying any non-zero number by its reciprocal is one (1).

9.2.0. Exchanging the numerator and denominator in a non-zero fraction produces the reciprocal of the original fraction.

9.3.0. Division tells one how many identical groups to separate a quantity into OR how many items to place in each group.

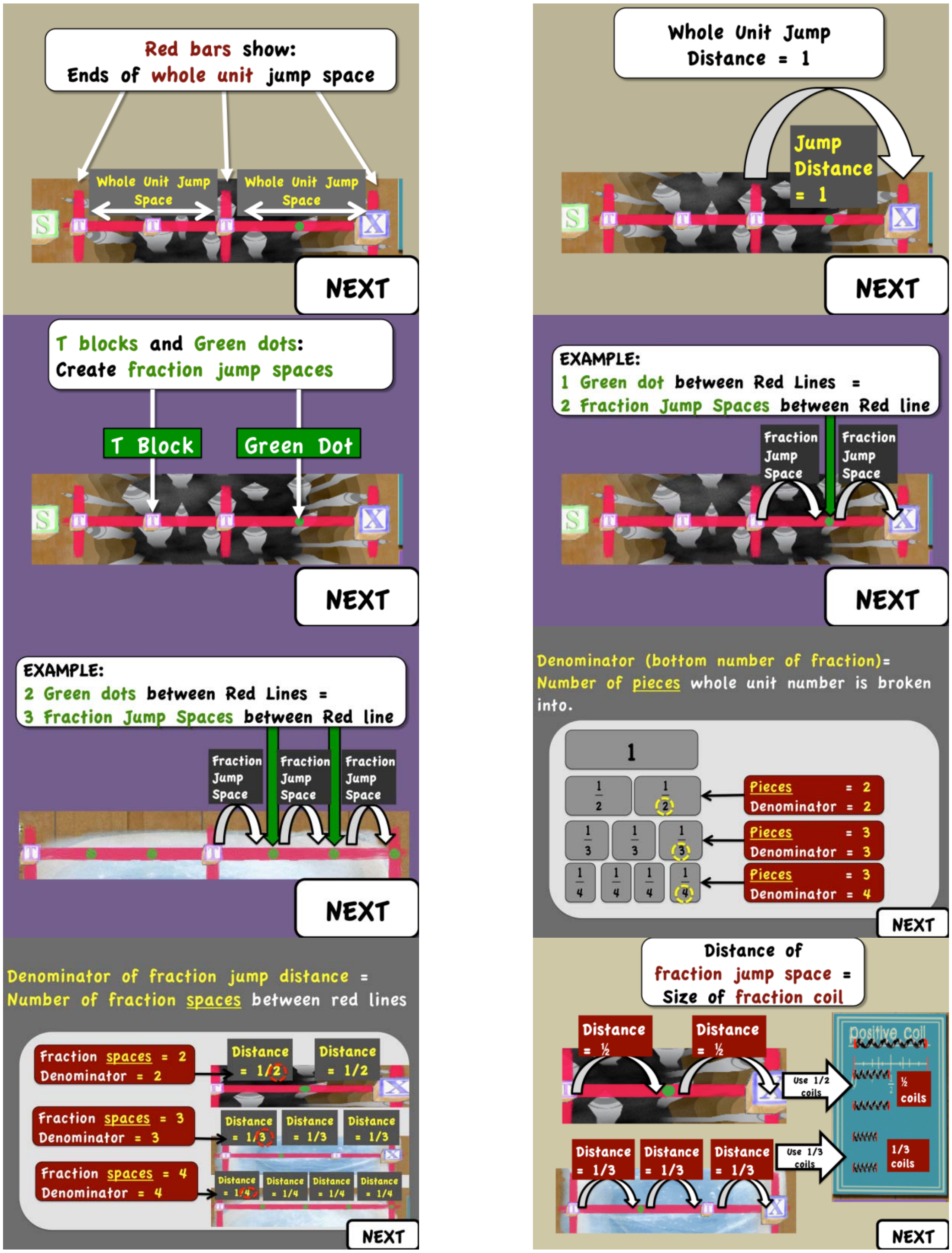
9.3.1. If separating a quantity into identical groups, the quotient of a division tells how many items are in each WHOLE group.

9.3.2. If placing a specific number of items in each group, the quotient of a division tells how many groups one can make.

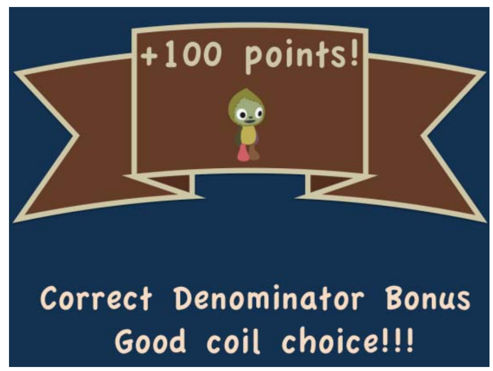
Appendix C: Scored Event Rationale

Scored event	Game knowledge required	Math knowledge required
Choosing the coil size	The vertical red bars denote the whole unit.	In mathematics, one unit is understood to be one of some quantity (intervals, areas, volumes, etc.). In our number system, the unit can be represented as one whole interval on a number line. Positive integers are represented by successive whole intervals on the positive side of zero. The interval between each integer is constant once it is established.
	Grid: The *spaces* between the green dots are the *parts* of the whole unit. Coil: The coil *pieces* are *parts* of a whole unit coil.	Positive non-integers are represented by fractional parts of the interval between whole numbers.
	Grid: The number of *spaces* between the green dots is the denominator. Coil: The number of coil *pieces* the whole unit is broken into is the denominator.	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.
Adding coils	If given different coils with different units, the coils must be changed so that they are the same unit before they can be added together.	Only identical (common) units can be added to create a single numerical sum.
Patch reaches the goal	The length of the jump is the number of pieces between the blocks.	Positive integers can be broken (decomposed) into parts that are each one unit in quantity.
	Add the correct number of coils that match the length of the jump.	All rational numbers can be represented as additions of integers or fractions. To add quantities, the units (or parts of units) must be identical. Identical (common) units can be added to create a single numerical sum.
	Grid: The top number of the jump distance equals the total number of spaces to jump over. Coils: The top number of the sum of the coil pieces on the trampoline represents the number of coil pieces that have been added together.	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.

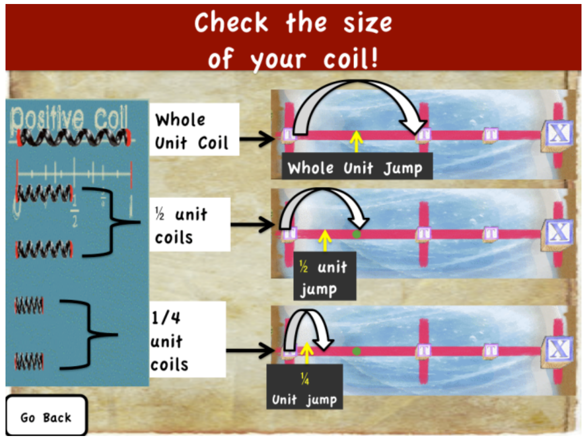
Appendix D:
Example Tutorial Images: Stage 2

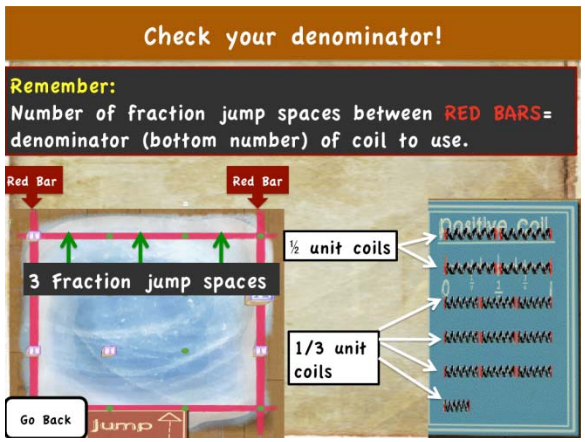


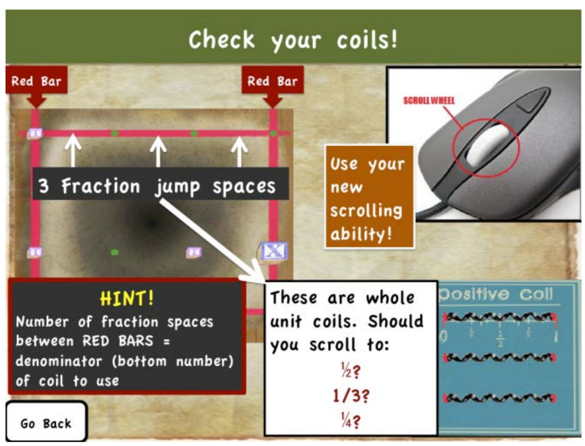
Appendix E:
Feedback Images



Appendix F:
Sample Additional Feedback







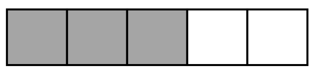
Appendix G:
Item Specifications (Sample)

Knowledge Specs	Computational Fluency: Students can execute procedures in the domain without the need to create or derive the procedure. Fluid performance is based on recall of patterns or other well established procedures, and is fast, automatic, and error-free. *How is something done?*		Conceptual Understanding: Captures demonstration of understanding of the mathematical concepts. *Why is something done?*	
	When presented with… *(Assessment Stimulus)*	Students should be able to…	When presented with… *(Assessment Stimulus)*	Students should be able to…
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.	Any rational number…	Place it on a number line relative to the whole interval explicitly (0 and 1 labeled) or implicitly (0 and an integer other than 1 labeled) defined.	Apparent contradictions involving rational number such as $\frac{3}{4} < \frac{1}{2}$ or $\frac{1}{2}$ does not equal $\frac{1}{2}$.	Explain that the contradiction can be resolved if their relative wholes must be equal when comparing.
	Given a unit whole (interval, volume, area, etc.)…	Show how much of the whole must be shaded to represent a fractional amount.		
1.2.0. In mathematics, one unit is understood to be one of some quantity (intervals, areas, volumes, etc.).	A histogram of a certain quantity represented by discrete objects…	Identity the unit that each single discrete object represents (e.g. each rose represents thousands of flowers sold on Valentine's Day).	Given a relationship between a real world measure and a scale model…	Explain how what size of unit to use on the model to accurately represent the real world quantity.

Appendix H:
Pretest Administered

1. Time started: ________________

2. Write a fraction that describes the shaded part of the figure below.



Answer: _________

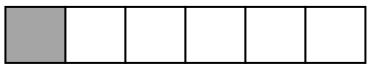
Fill in the boxwith a number that will make the statement true.

3. $$\frac{2}{10} + \frac{4}{10} + \frac{1}{10} = \frac{7}{\square}$$

4. A student has finished **four pages** of a six-page test. Write a fraction that shows the part of the test the student has completed.

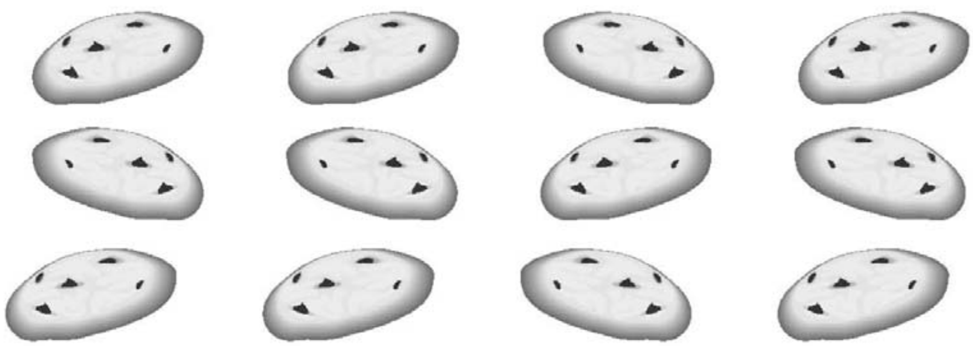
Answer: ____________________

5. In the figure below, **how many MORE** small squares need to be shaded so that $\frac{5}{6}$ of the small squares are shaded?



Answer: _________

6. There are 12 cookies. **Draw a circle** around $\frac{1}{3}$ of the cookies.



For the questions below, fill in each box with a number that will make the statement true.
The fractions DO NOT need to be simplified!

7. $\frac{1}{5} + \frac{2}{5} = \frac{\square}{\square}$

The fraction does not need to be simplified.

8. $\frac{1}{6} + \frac{\square}{\square} = \frac{5}{6}$

The fraction does not need to be simplified.

Use this information to fill in the boxes below:

$$\frac{1}{2} \text{ is the same as } \frac{2}{4} \text{ and } \frac{3}{6} \text{ and } \frac{4}{8} \text{ and } \frac{5}{10}$$

9. $\frac{1}{2} + \frac{3}{4} = \frac{\square}{4}$

10. $\frac{1}{2} + \frac{4}{10} = \frac{\square}{10}$

11. $\frac{1}{2} + \frac{3}{6} = \frac{6}{\square}$

12. $\frac{1}{2} + \frac{8}{8} = \frac{12}{\square}$

13. How many $\frac{1}{4}$'s are in $\frac{3}{4}$?

a. Answer: _________

b. What does the numerator of 3 tell you in $\frac{3}{4}$? **Choose only one answer.**

   a. It tells you there are three $\frac{1}{4}$'s in this fraction
   b. It tells you the whole unit is broken into three pieces
   c. It tells you there are three whole units in this fraction
   d. It tells you to add 4 three times

c. What does the denominator of 4 tell you in $\frac{3}{4}$? **Choose only one answer.**

   a. It tells you there are four $\frac{3}{4}$'s in this fraction
   b. It tells you the whole unit is broken into four pieces
   c. It tells you there are four whole units in this fraction
   d. It tells you to add 3 four times

14. Using the picture above, choose the answer that **best** explains why the denominator is "2" in $\frac{1}{2}$

   a. There are 2 whole units.

   b. You multiplied the "1" by 2.

   c. The number "1" has been divided into two pieces.

   d. Because 1 + 1 = 2.

Fill in **EACH** box with a number that will make the statement true.

15. $$\frac{1}{3} + \frac{\square}{3} + \frac{\square}{3} + \frac{\square}{3} = \frac{4}{3}$$

16. Choose the fraction that represents the shaded boxes.

a. $\frac{4}{2}$

b. $\frac{2}{4}$

c. $\frac{1}{4}$

d. $\frac{2}{8}$

17. Choose the answer that best explains what the numerator of a fraction is.

a. How many equal parts you have.

b. How many whole units you have in a fraction.

c. How many equal parts you broke the whole unit into.

For the problems below, indicate whether or not they were completed correctly or not. If you do not know, mark "Don't Know."

	Correct	Incorrect	Don't Know
18. $\frac{1}{8}+\frac{2}{4}=\frac{3}{12}$	☐	☐	☐
19. $\frac{3}{4}+\frac{3}{2}=\frac{9}{4}$	☐	☐	☐
20. $\frac{5}{6}+\frac{2}{6}=\frac{7}{6}$	☐	☐	☐
21. $\frac{3}{6}+\frac{5}{6}=\frac{8}{6}$	☐	☐	☐
22. $\frac{2}{11}+\frac{2}{11}=\frac{4}{11}$	☐	☐	☐
23. $\frac{1}{2}+\frac{2}{4}=\frac{3}{6}$	☐	☐	☐
24. $\frac{1}{3}+\frac{3}{6}=\frac{5}{6}$	☐	☐	☐

25. How would you explain how to solve this problem?

$$\frac{2}{11}+\frac{2}{11}=\frac{4}{11}$$

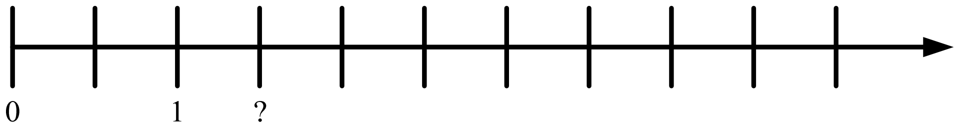
26. How would you explain how to solve this problem?

$$\frac{1}{2}+\frac{2}{4}=\frac{3}{6}$$

27. How would you explain how to solve this problem?

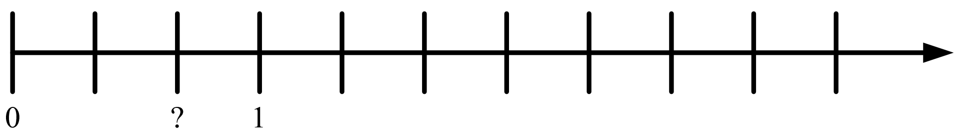
$$\frac{1}{3}+\frac{3}{6}=\frac{5}{6}$$

28. At what number is the “?” located?



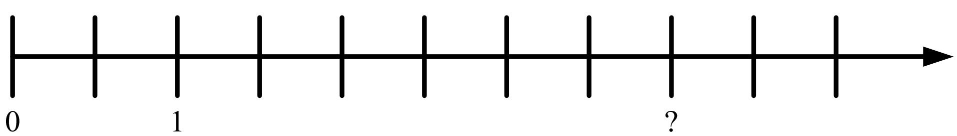
Answer: _________

29. At what number is the “?” located?



Answer: _________

30. At what number is the “?” located?



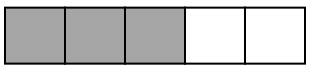
Answer: _________

31. Time finished: ________________

Appendix I:
Administered Posttest

1. Time started: ________________

2. Write a fraction that describes the shaded part of the figure below.



Answer: _________

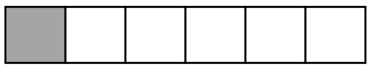
Fill in the box with a number that will make the statement true.

3. $$\frac{2}{10} + \frac{4}{10} + \frac{1}{10} = \frac{7}{\square}$$

4. A student has finished **four pages** of a six-page test. Write a fraction that shows the part of the test the student has completed.

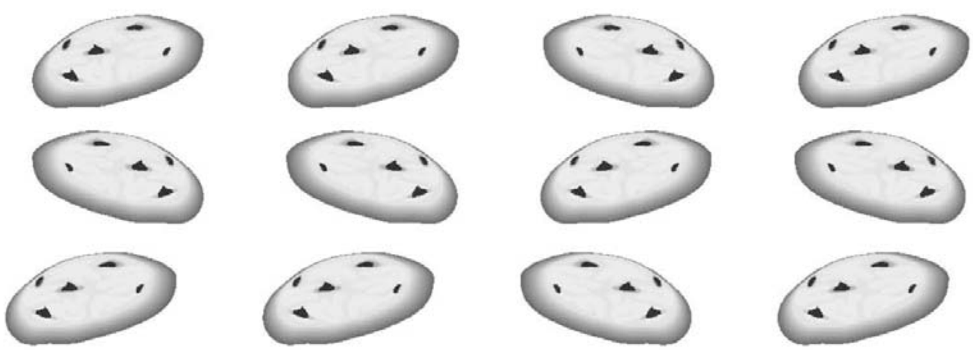
Answer: ____________________

5. In the figure below, **how many MORE** small squares need to be shaded so that $\frac{5}{6}$ of the small squares are shaded?



Answer: _________

6. There are 12 cookies. **Draw a circle** around $\frac{1}{3}$ of the cookies.



For the questions below, fill in each box with a number that will make the statement true.
The fractions DO NOT need to be simplified!

7. $\frac{1}{5} + \frac{2}{5} = \frac{\square}{\square}$

The fraction does not need to be simplified.

8. $\frac{1}{6} + \frac{\square}{\square} = \frac{5}{6}$

The fraction does not need to be simplified.

Use this information to fill in the boxes below:

$$\frac{1}{2} \text{ is the same as } \frac{2}{4} \text{ and } \frac{3}{6} \text{ and } \frac{4}{8} \text{ and } \frac{5}{10}$$

9. $\dfrac{1}{2} + \dfrac{3}{4} = \dfrac{\square}{4}$

10. $\dfrac{1}{2} + \dfrac{4}{10} = \dfrac{\square}{10}$

11. $\dfrac{1}{2} + \dfrac{3}{6} = \dfrac{6}{\square}$

12. $\dfrac{1}{2} + \dfrac{8}{8} = \dfrac{12}{\square}$

13. How many $\frac{1}{4}$'s are in $\frac{3}{4}$?

a. Answer: _________

b. What does the numerator of 3 tell you in $\frac{3}{4}$? **Choose only one answer.**

   a. It tells you there are three $\frac{1}{4}$'s in this fraction

   b. It tells you the whole unit is broken into three pieces

   c. It tells you there are three whole units in this fraction

   d. It tells you to add 4 three times

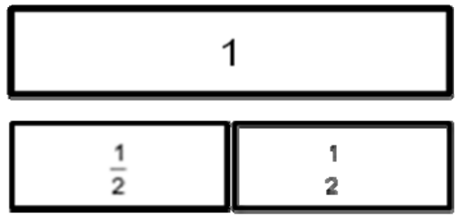
c. What does the denominator of 4 tell you in $\frac{3}{4}$? **Choose only one answer.**

   a. It tells you there are four $\frac{3}{4}$'s in this fraction

   b. It tells you the whole unit is broken into four pieces

   c. It tells you there are four whole units in this fraction

   d. It tells you to add 3 four times



14. Using the picture above, choose the answer that **best** explains why the denominator is “2” in $\frac{1}{2}$

a. There are 2 whole units.

b. You multiplied the “1” by 2.

c. The number “1” has been divided into two pieces.

d. Because 1 + 1 = 2.

Fill in **EACH** box with a number that will make the statement true.

15. $$\frac{1}{3} + \frac{\square}{3} + \frac{\square}{3} + \frac{\square}{3} = \frac{4}{3}$$

16. Choose the fraction that represents the shaded boxes.

a. $\frac{4}{2}$

b. $\frac{2}{4}$

c. $\frac{1}{4}$

d. $\frac{2}{8}$

17. Choose the answer that best explains what the numerator of a fraction is.

a. How many equal parts you have.

b. How many whole units you have in a fraction.

c. How many equal parts you broke the whole unit into.

For the problems below, indicate whether or not they were completed correctly or not. If you do not know, mark "Don't Know."

	Correct	**Incorrect**	**Don't Know**
18. $\frac{1}{8}+\frac{2}{4}=\frac{3}{12}$	☐	☐	☐
19. $\frac{3}{4}+\frac{3}{2}=\frac{9}{4}$	☐	☐	☐
20. $\frac{5}{6}+\frac{2}{6}=\frac{7}{6}$	☐	☐	☐
21. $\frac{3}{6}+\frac{5}{6}=\frac{8}{6}$	☐	☐	☐
22. $\frac{2}{11}+\frac{2}{11}=\frac{4}{11}$	☐	☐	☐
23. $\frac{1}{2}+\frac{2}{4}=\frac{3}{6}$	☐	☐	☐
24. $\frac{1}{3}+\frac{3}{6}=\frac{5}{6}$	☐	☐	☐

25. How would you explain how to solve this problem?

$$\frac{2}{11}+\frac{2}{11}=\frac{4}{11}$$

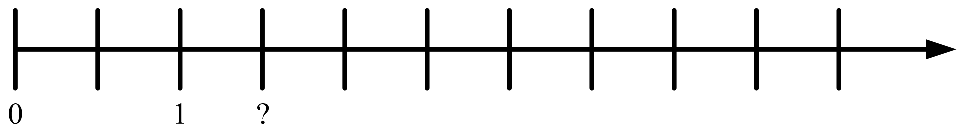
26. How would you explain how to solve this problem?

$$\frac{1}{2}+\frac{2}{4}=\frac{3}{6}$$

27. How would you explain how to solve this problem?

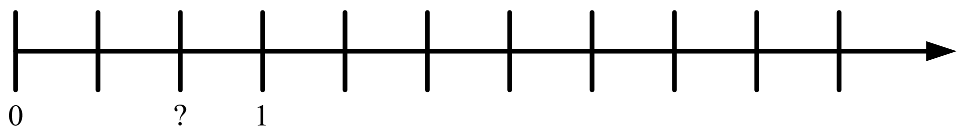
$$\frac{1}{3}+\frac{3}{6}=\frac{5}{6}$$

28. At what number is the “?” located?



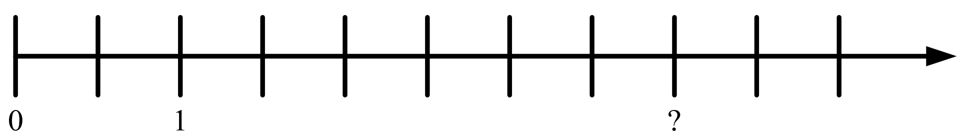
Answer: _________

29. At what number is the “?” located?



Answer: _________

30. At what number is the “?” located?



Answer: _________

31. You are given a $\frac{1}{5}$ coil and a $\frac{2}{5}$ coil. How far will Patch bounce if you add both of those coils to one trampoline?

Answer: _________

32. You are given a $\frac{1}{6}$ coil and a $\frac{1}{4}$ coil. If you want to use both of them to make Patch jump, what would you need to do to make sure you could add these two coils to a single trampoline **in the game**? Choose the answer that **best** explains.

a. Add 6 + 4 and then 1 + 1

b. Multiply 6 × 1 and 4 × 1

c. Change the fraction to 12ths

d. Change the $\frac{1}{6}$ into $\frac{1}{4}$

33. You are given a $\frac{1}{6}$ coil and a $\frac{1}{4}$ coil. How far will Patch bounce after you add both of those coils to one trampoline?

Answer: _________

34. **Select ALL the answers** that explain, mathematically, why you cannot combine different size coils in the game.

☐ You cannot add unlike pieces together

☐ The game just won't let you add certain things together

☐ The numerators have to be the same

☐ Addition tells you how many identical things you have

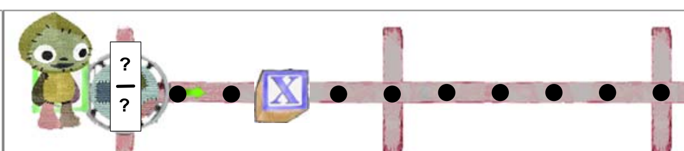
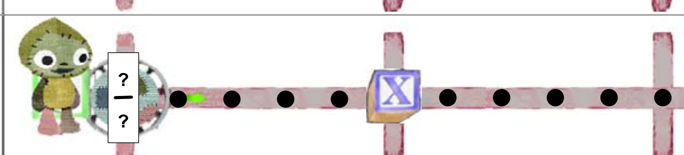
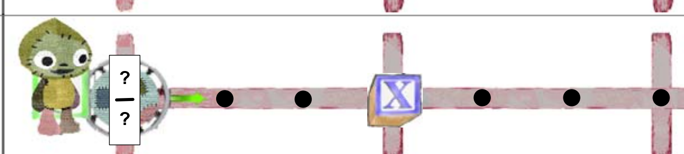
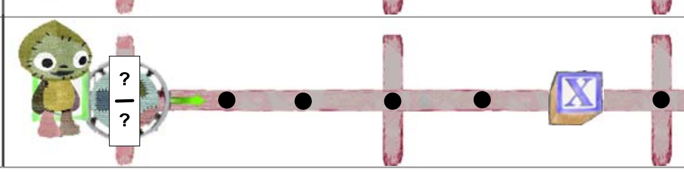
Suppose a trampoline had a $\frac{3}{5}$ coil on it.

Given the situation above, for each statement below, circlewhether the statement is true (T) or false (F).

35.	T	F	The unit should be divided into 3 equal pieces.	The unit is the distance between the vertical lines.
36.	T	F	The unit should be divided into 5 equal pieces.	
37.	T	F	If Patch jumped on the trampoline, he would bounce 3 pieces.	
38.	T	F	If Patch jumped on the trampoline, he would bounce 5 pieces.	

Suppose a trampoline with a $\frac{3}{2}$ coil was placed on the grid as shown below.

39. Place an "X" on the spot where Patch would land after pressing *Jump*.

40. What coil value will make Patch bounce to the X block?	$\frac{\square}{\square}$	
41. What coil value will make Patch bounce to the X block?	$\frac{\square}{\square}$	
42. What coil value will make Patch bounce to the X block?	$\frac{\square}{\square}$	
43. What coil value will make Patch bounce to the X block?	$\frac{\square}{\square}$	

Appendix J:
Administered Post-Game Survey

For each question, circle the number that shows how much you agree with the statement.

OVERALL COMMENTS ON THE GAME

Indicate how much you AGREE with the following statements.	Disagree	Disagree a little	Agree a little	Agree
1. I knew how I lost points	1	2	3	4
2. I knew how I scored points	1	2	3	4
3. I thought the directions made sense	1	2	3	4
4. I think the game helped make my math skills better	1	2	3	4
5. I was confused about how to play the game	1	2	3	4
6. I learned something new about fractions from the game	1	2	3	4
7. If someone asked me, I could explain what the green dots mean	1	2	3	4
8. I cared about earning points in the game	1	2	3	4
9. I was concentrating a lot while playing the game	1	2	3	4
10. I became hooked on the game	1	2	3	4
11. It felt like I was playing the game for less time than I really did	1	2	3	4
12. I enjoyed playing the game	1	2	3	4
13. I would play this game again	1	2	3	4
14. I forgot about everything else around me while I was playing the game	1	2	3	4
15. I was not able to pay attention to the game	1	2	3	4
16. I was paying attention to other things while I played the game	1	2	3	4
17. Beating the different levels made me feel good	1	2	3	4
18. I really got into the game	1	2	3	4
19. Playing the game was boring	1	2	3	4
20. If the game had more levels, I would want to play them	1	2	3	4
21. I would have liked to play longer	1	2	3	4
22. I got annoyed playing the game	1	2	3	4
23. I had to try really hard while playing the game	1	2	3	4
24. I thought the game looked cool	1	2	3	4
25. I did not want to lose in the game	1	2	3	4
26. I thought the game was fun	1	2	3	4
27. I thought the game was frustrating	1	2	3	4
28. Time seemed to go by very quickly when I played the game	1	2	3	4
29. I liked that the game was hard sometimes	1	2	3	4
30. I wish I had more time to play the game	1	2	3	4
31. I would play this game when I have free time	1	2	3	4
32. I thought the game was hard	1	2	3	4
33. I lost track of time when playing the game	1	2	3	4
34. I would tell my friends to play this game	1	2	3	4
35. The game is like other puzzle games I play	1	2	3	4
36. I knew what the goals of the game were	1	2	3	4
37. This game was as much fun as other puzzle games I enjoy	1	2	3	4
38. I can see similarities between this game and other puzzle games	1	2	3	4

Indicate how much you AGREE with the following statements.	Disagree	Disagree a little	Agree a little	Agree
39. This game is similar to other puzzle games	1	2	3	4

Indicate how FREQUENTLY the following occurred …	Almost never	Sometimes	Often	Almost always
40. I read the directions	1	2	3	4
41. I just looked at how big the coil was (and not its value, like 1/2 or 1/3) to choose what size coil to use	1	2	3	4
42. I guessed when I chose what size coil to use	1	2	3	4
43. I used math to figure out what size coil to use	1	2	3	4
44. I used the Reset button	1	2	3	4
45. I used the Reset Tutorial (go back to tutorial) button	1	2	3	4
46. I used Patch's Survival Guide	1	2	3	4

47. Please write down any comments you have about the game.

HOW OFTEN YOU PLAY GAMES

Circle the number that shows how often you play the following types of games in general *(note: if you don't know what a term means, mark the first box).*

Indicate how much you play the following types of games.	Don't know the word	Hardly ever	Some-times	Often	Very often
48. Puzzle (ex: Tetris, Minesweeper, Bejeweled)	☐	1	2	3	4
49. Real Time Strategy (RTS; ex: Age of Empires, Command & Conquer)	☐	1	2	3	4
50. Action (ex: Halo, SOCOM)	☐	1	2	3	4
51. Role Playing (ex:Neverwinter Nights, World of Warcraft)	☐	1	2	3	4
52. Sports (ex: Madden Football, Tiger Woods Golf)	☐	1	2	3	4
53. First-person perspective or shooter (ex: Halo, Call of Duty, Resistance)	☐	1	2	3	4
54. Arcade style (ex: Pac-man, Pong)	☐	1	2	3	4
55. Console games (ex: Xbox, Playstation, Wii, DS, PSP, Gamecube)	☐	1	2	3	4
56. Mobile/phone games (ex:iPhone, iTouch, Blackberry, Android, Palm Pre)	☐	1	2	3	4

VIDEO GAME AND COMPUTER EXPERIENCE

57. How many HOURS a WEEK do you play video games (computer, console, handheld)? (Estimate if you don't know, or think about how many hours a day you play and add them all up for the week!)

☐0 hours/week ☐1-4 hours/week ☐5-8 hours/week ☐9-12 hours/week ☐13+ hours/week

58. How would you describe your skill level with video games?

☐Poor ☐Fair ☐Good ☐Very good

59. How often do you use a computer? (not including video games):___________ hours per week

60. How would you describe your skill level with computers?

☐Poor ☐Fair ☐Good ☐Very good

ATTITUDES ABOUT MATH

A number of statements which people have used to describe themselves are given below. For each statement, find the word or phrase which best describes how you think or feel and circle the number for your answer. There are no right or wrong answers. Do not spend too much time on any one statement.

Indicate how much the following occurs in general.	**Almost never**	**Some-times**	**Often**	**Almost always**
61. I know I can understand most information in a textbook even if it's difficult	1	2	3	4
62. If I want to learn something well, I can.	1	2	3	4
63. I'm confident I can do an excellent job on assignments and tests.	1	2	3	4
64. If I decide not to get any bad grades, I can really do it.	1	2	3	4
65. When I sit myself down to learn something really difficult, I can learn it.	1	2	3	4
66. I know that I can do well on the things I'm taught	1	2	3	4
67. If I decide not to get any problems wrong, I can really do it.	1	2	3	4
68. I know that I can understand even the hardest things taught by my teacher.	1	2	3	4

Indicate how much you AGREE with the following statements.	**Disagree**	**Disagree a little**	**Agree a little**	**Agree**
69. I learn faster if I'm trying to do better than others.	1	2	3	4
70. It is helpful to put together everyone's ideas when working on a project.	1	2	3	4
71. Mathematics is important to me personally.	1	2	3	4
72. I like to help other people do well in a group.	1	2	3	4
73. I would like to be the best at something.	1	2	3	4
74. Mathematics is one of my best subjects.	1	2	3	4
75. Trying to be better than others makes me work well.	1	2	3	4
76. Because doing mathematics is fun, I wouldn't want to give it up.	1	2	3	4
77. I get good grades in mathematics.	1	2	3	4
78. I like to try to be better than other students.	1	2	3	4
79. I like to work with other students.	1	2	3	4
80. When I work on math, I sometimes am so focused I forget how long I've been working.	1	2	3	4
81. I learn most when I work with other students.	1	2	3	4
82. I do my best work when I work with other students.	1	2	3	4
83. I have always done well in mathematics.	1	2	3	4

BACKGROUND

84. Birth date: ________ /________
Month Year

85. Grade:☐4th☐ 5th☐ 6th☐ 7th☐ 8th☐ 9th☐ 10th☐ 11th☐ 12th

86. What are you learning in your math class now?
____

87. Gender: ☐ Male ☐ Female

88. Ethnicity (choose only one):
- ☐ Biracial/multiethnic
- ☐ African-American
- ☐ Asian or Pacific Islander
- ☐ Hispanic / Latino/a
- ☐ Native-American
- ☐ White, non-Hispanic
- ☐ Other ____________________

89. How often do people in your home talk to each other in a language other than English?

☐Never ☐Once in a while ☐About half of the time ☐All or most of the time

90. What was your math grade on your last report card?

☐A ☐B ☐C ☐D ☐F ☐Don't know

91. What were your math grades last year?

☐A ☐B ☐C ☐D ☐F ☐Don't know

92. Did you play a version of this game before? ☐ Yes ☐ No

Appendix K:
Rubric for Open-Ended Items

How would you explain how to solve this problem?

$$\frac{2}{11}+\frac{2}{11}=\frac{4}{11}$$

Score Point	Description	Example
0	*Add the denominator* If a student mentions that the denominators should be added (i.e., using words or symbolically, or graphically), then automatically scored incorrect.	- "Add the top and bottom numbers" - Only rewrites the problem
0	*Unrelated response* Student mentions some non-mathematical strategy (e.g., something that is metacognitive) or judgment (e.g., that it was right/wrong).	- "Look at it" - "Check it" - "Correct"
1	*Add fractions* Response indicates that the fractions can be added together but does not specifically mention that only the numerators should be added.	- "Add the fractions together"
2	*Add numerators* If student mentions that numerators should be added, (i.e., using words or symbolically, or graphically), then scored correct.	- Only has "2+2 = 4" - "Add 2 and 2" - "All you need to do is add the numerator" - "Just add the top and leave the bottom number the same"
2	*Represents the right answer* Student uses a diagram or picture to demonstrate correctly how the problem should be solved	- Has a diagram with a rectangle split into 11 pieces and four are shaded
98	Don't know	- Says something like "DK" or "?"
99	Missing	- No response
N/A	Cannot interpret	

How would you explain how to solve this problem?

$$\frac{1}{2}+\frac{2}{4}=\frac{3}{6}$$

Score Point	Description	*Example*
0	*Add without changing anything* Student mentions adding the (a) fractions, **OR** (b) numerators, **OR** (c) denominators, **OR** (d) numbers without changing or converting the numbers.	- "Add the top and bottom numbers" - Says "1+2 = 3" and "Add 2+4=6" - "Add the numerators"
0	*Cannot add denominators* Response does not indicate that anything should be changed but does acknowledge that the denominators cannot be added together.	- "Do not add denominators" - "Keep bottom numbers the same"
0	*Unrelated response* Student mentions some non-mathematical strategy (e.g., something that is metacognitive) or judgment (e.g., that it was right/wrong).	- "Look at it" - "Check it" - "Correct"
0	*Wrong response* Response is completely wrong (does not indicate that the student understands that fractions need to be converted).	- "Are not equal" - "Divide"
1	*Mentions converting the fraction or that the denominator needs to be changed* Student's response refers to something having to be changed (e.g., (a) fractions, **OR** (b) numerators, **OR** (c) denominators, **OR** (d) numbers) but **does not** mention the need for a common denominator. **partial credit is given for acknowledging that something needs to be changed**	- "Change the denominators" - "Change 1 of the fractions and then add it" - "You need to convert the fractions"
1	*Mentions converting the fraction or that the denominator needs to be changed but inaccurate computation* Student's response refers to something having to be changed (e.g., (a) fractions, **OR** (b) numerators, **OR** (c) denominators, **OR** (d) numbers) but **does an inaccurate computation.** **partial credit is given for acknowledging that something needs to be changed**	- "Change the denominator into an 8" - "½ = 2/6"
1	*Mentions finding a common denominator but no mention of converting entire fraction* Student's response refers to the need to find a common denominator but does not refer to needing to then convert the fraction. **partial credit is given for acknowledging that a common denominator is needed but does not get full credit because student does not acknowledge that the entire fraction needs to be converted***	- "Find the LCD and then add them"

Score Point	Description	*Example*
2	*Convert fractions to a common denominator* Response indicates that the fraction should be converted to have a common denominator before they can be added.	- "You can't solve it unless you have equal parts" - "You need to convert the fractions into a common denominator" - "Multiply the top and bottom number by 2"
2	*Graphically or symbolically represents the process or product of the addition* Using symbols or diagrams, shows how the problem should be solved or shows the conversion.	- "2/4+2/4 = 4/4" - "½ = 2/4" - Shows two circles with one of them ½ shaded, and the other broken into four parts with two shaded
98	Don't know	- Says something like "DK" or "?"
99	Missing	- No response
N/A	Cannot interpret	

How would you explain how to solve this problem?

$$\frac{1}{3}+\frac{3}{6}=\frac{5}{6}$$

Score Point	Description	Example
0	*Add without changing anything* Student mentions adding the (a) fractions, **OR** (b) numerators, **OR** (c) denominators, **OR** (d) numbers without changing or converting the numbers.	- "Add the top and bottom numbers" - "1+3" - "Add 3+6" - "Add the numerators" - "Adding the numbers" - "6 is common so add the numerators" - "1/3 + 3/6=4/9"
0	*Cannot add denominators* Response does not indicate that anything should be changed but does acknowledge that the denominators cannot be added together.	- "Do not add denominators" - "Keep bottom numbers the same"
0	*Unrelated response* Student mentions some non-mathematical strategy (e.g., something that is metacognitive) or judgment (e.g., that it was right/wrong)	- "Look at it" - "Check it" - "Wrong"
0	*Wrong response* Response is completely wrong (does not indicate that the student understands that fractions need to be converted)	- "Are not equal" - "Divide"
1	*Mentions converting the fraction or that the denominator needs to be changed* Student's response refers to something having to be changed [e.g.,(a) fractions, **OR** (b) numerators, **OR** (c) denominators, **OR** (d) numbers] but **does not** mention the need for a common denominator **partial credit is given for acknowledging that something needs to be changed**	- "Change the denominators" - "Change 1 of the fractions and then add it" - "You need to convert the fractions" - "It needs to be changed"
1	*Mentions converting the fraction or that the denominator needs to be changed but inaccurate computation* Student's response refers to something having to be changed (e.g., (a) fractions, **OR** (b) numerators, **OR** (c) denominators, **OR** (d) numbers) but **does an inaccurate computation.** **partial credit is given for acknowledging that something needs to be changed**	- "Change the 1/3 to 1/6"
1	*Mentions finding a common denominator but no mention of converting entire fraction* Student's response refers to the need to find a common denominator but does not refer to needing to then convert the fraction. **partial credit is given for acknowledging that a common denominator is needed but does not get full credit because student does not acknowledge that the entire fraction needs to be converted***	- "Find the LCD and then add the numerators"

Score Point	Description	Example
2	*Convert fractions to a common denominator* Response indicates that the fraction should be converted to have a common denominator before they can be added.	- "You can't solve it unless you have equal parts" - "You need to convert the fractions into a common denominator"
98	*Graphically or symbolically represents the process or product of the addition* Using symbols or diagrams, shows how the problem should be solved or shows the conversion.	- Shows two circles with one of them 1/3 shaded, and the other broken into six parts with three shaded - "1/3= 2/6"
98	Don't know	- Says something like "DK" or "?"
99	Missing	- No response
N/A	Cannot interpret	- Illegible

Appendix L:
Rationale for Intraclass Correlation Coefficient Model

Reliability for each item was determined by calculating the Intraclass Correlation Coefficients (ICC) for each item. Choosing the appropriate ICC model is determined by the nature of the data and what is examined to be reliable (McGraw & Wong, 1996). A two-way, mixed-effect model was used to examine the absolute agreement of measurements between the three raters for each of the three open-ended items. In these data, the rows represent measurements for different participants. The measurement observations are independent (i.e., for each item, data are not coming from the same individual more than once). Thus, these measurements are one systematic source of variance. The columns represent different raters, which were another systematic source of variance. When there are two sources of variance, a two-way model is used. In these data, the ICCs were applied to single measurement scores (i.e., individual ratings of items by individual judges); therefore, an ICC model that applies to single measurements was used. For the purposes of the study, the agreement among raters, rather than the consistency within a rater was of utmost concern. Thus, an absolute agreement of correlation is used. Finally, it could not be assumed that findings from the calculation of these three raters will generalize to other raters; thus, the column variable, raters, was considered a fixed variable.

Appendix M:
Descriptive Statistics for Background Variables

Table M1

Descriptive Statistics on Frequency of Game Play by Game Type

Game type	*M*	Hardly ever	Sometimes	Often	Very often
Puzzle	2.57	14 (10.93%)	28 (21.87%)	8 (6.25%)	24 (18.75%)
Real Time Strategy	2.26	30 (23.43%)	15 (11.71%)	7 (5.46%)	21 (16.4%)
Action	2.97	10 (7.81%)	18 (14.06%)	15 (11.71%)	36 (28.12%)
Role Playing	2.41	29 (22.65%)	15 (11.71%)	9 (7.03%)	26 (20.31%)
Sports	2.84	16 (12.5%)	16 (12.5%)	19 (14.84%)	34 (26.56%)
First-person perspective or shooter	2.98	18 (14.06%)	10 (7.81%)	11 (8.59%)	44 (34.37%)
Arcade style	3.00	7 (5.46%)	25 (19.53%)	17 (13.28%)	39 (30.46%)
Console games	3.49	6 (4.68%)	9 (7.03%)	9 (7.03%)	64 (50%)
Mobile/phone games	3.06	17 (13.28%)	11 (8.59%)	9 (7.03%)	50 (39.06%)

Table M2

Descriptive Statistics: No. of Hours Video Games Played; Video Game Skill Level; Computer Skill Level

Video game and computer experience	*n*	%
Number of hours video games are played		
0 hours/week	7	5.5
1-4 hours/week	58	45.3
5-8 hours/week	20	15.6
9-12 hours/week	5	3.9
13+ hours/week	12	9.4
Video game skill level		
Poor	4	3.1
Fair	13	10.2
Good	27	21.1
Very good	61	47.7
Computer skill level		
Poor	4	3.1
Fair	13	10.2
Good	27	21.1
Very good	61	47.7

Table M3

Descriptive Statistics: Self-Efficacy and Learning Style Preferences

Cognitive factor	*n*	Min.	Max.	*M*	*SD*	Variance
Self-efficacy	101	1.38	4.00	3.22	0.68	0.46
Math self-concept	98	1.50	4.00	3.28	0.66	0.44
Competitive learning	98	1.00	4.00	3.13	0.77	0.59
Cooperative learning	98	1.60	4.00	3.29	0.72	0.52

Table M4

ANOVAs for Background Variables

Background variable	*SS*	*df*	Mean Square	*F*	*p*
Gaming experience					
Between groups	1.268	2	.634	1.050	.355
Within groups	48.913	81	.604		
Total	50.181	83			
Self-efficacy					
Between groups	.278	2	.139	.298	.743
Within groups	39.553	85	.465		
Total	39.830	87			
Math self-concept					
Between groups	.052	2	.026	.058	.944
Within groups	37.363	83	.450		
Total	37.415	85			
Competitive					
Between groups	.819	2	.409	.678	.511
Within groups	50.142	83	.604		
Total	50.961	85			
Cooperative					
Between groups	1.428	2	.714	1.288	.281
Within groups	46.020	83	.554		
Total	47.449	85			

Appendix N:
Correlation Matrix: Game Performance and Math Achievement Measures

Table N1

Correlation Matrix: Game Performance and Math Achievement Measures

Measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Pretest	1.00	0.90**	0.55**	-0.12	0.10	0.10	0.29**	-0.08	-0.18	0.06	0.26*	0.20	0.41**	0.29**
Posttest	0.90**	1.00	0.61**	0.32**	0.02	0.11	0.26**	-0.11	-0.21*	0.00	0.16	0.13	0.33**	0.25*
Game context items	0.55**	0.61**	1.00	0.22*	-0.02	0.09	0.16	-0.26**	-0.22*	0.02	0.06	-0.06	0.19	0.09
Normalized changes scores	-0.12	0.32**	0.22*	1.00	-0.19	0.04	-0.02	-0.06	-0.05	-0.15	-0.18	-0.11	-0.07	0.01
Added coils unlike denominators	0.10	0.02	-0.02	-0.19	1.00	0.20*	0.40**	0.33**	0.44**	0.21*	0.01	-0.09	0.03	-0.08
Using wrong sized coils	0.10	0.11	0.09	0.04	0.20*	1.00	0.68**	0.17	0.15	-0.01	0.01	-0.03	0.04	0.06
Adding coils	0.29**	0.26**	0.16	-0.02	0.40**	0.68**	1.00	0.31**	0.33**	0.10	0.12	0.06	0.15	0.11
Resets	-0.08	-0.11	-0.26**	-0.06	0.33**	0.17	0.31**	1.00	0.74**	-0.01	0.00	-0.03	-0.08	-0.13
Failed attempts	-0.18	-0.21*	-0.22*	-0.05	0.44**	0.15	0.33**	0.74**	1.00	-0.01	-0.05	-0.03	-0.14	-0.17
Gaming experience	0.06	0.00	0.02	-0.15	0.21*	-0.01	0.10	-0.01	-0.01	1.00	0.30**	0.29**	0.36**	0.35**
Competitive	0.26*	0.16	0.06	-0.18	0.01	0.01	0.12	0.00	-0.05	0.30**	1.00	0.58**	0.56**	0.53**
Cooperative	0.20	0.13	-0.06	-0.11	-0.09	-0.03	0.06	-0.03	-0.03	0.29**	0.58**	1.00	0.61**	0.66**
Self-efficacy	0.41**	0.33**	0.19	-0.07	0.03	0.04	0.15	-0.08	-0.14	0.34**	0.54**	0.61**	1.00	0.69**
Math self-concept	0.29**	0.25*	0.09	0.01	-0.08	0.06	0.12	-0.13	-0.17	0.35**	0.53**	0.66**	0.69**	1.00