Journal of Research on Technology in Education

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ujrt20

The Effectiveness of Reason Racer, a Game Designed to Engage Middle School Students in Scientific Argumentation

Marilyn Ault[a], Jana Craig-Hare[a], Bruce Frey[b], James D. Ellis[b] & Janis Bulgren[c]

[a] ALTEC at the University of Kansas Center for Research on Learning

[b] University of Kansas School of Education

[c] University of Kansas Center for Research on Learning

Published online: 08 Jan 2015.

JRTE | Vol. 47, No. 1, pp. 21–40 | © 2015 ISTE | iste.org/jrte
DOI: 10.1080/15391523.2015.967542

The Effectiveness of Reason Racer, a Game Designed to Engage Middle School Students in Scientific Argumentation

Marilyn Ault and Jana Craig-Hare
ALTEC at the University of Kansas Center for Research on Learning

Bruce Frey and James D. Ellis
University of Kansas School of Education

Janis Bulgren
University of Kansas Center for Research on Learning

Abstract

Reason Racer is an online, rate-based, multiplayer game that applies specific game features in order to engage middle school students in introductory knowledge of and thinking related to scientific argumentation. Game features include rapid and competitive play, timed performance, immediate feedback, and high rates of response across many game-play sessions and science scenarios. The areas of argumentation addressed in the game include understanding a claim, judging evidence about a claim based on type (fact, opinion) and quality, determining the reasoning applied to the claim (authority, theory, or logic), considering counterarguments and rebuttals, and making judgments. These skills have been identified as important by previous research. Students who played the game at least 10 times improved in every aspect of argumentation skill and judgment. Students who played the game also reported an increase in confidence and motivation to engage in science compared to students who did not play the game. This study has implications for the use of rate-based, multiplayer, competitive games as a component of instruction for difficult-to-teach skills. We also assume that the engaging and fun aspects of Reason Racer contributed to students reporting an increased interest in science. (Keywords: educational game, scientific argumentation, middle school science, multiplayer game)

Argumentation is a very difficult skill to learn, yet it is essential for daily discourse and is a significant part of higher-order thinking and reasoning across curriculum content areas. The ability to reason has many manifestations, and one important aspect is the ability to evaluate claims or statements, especially when a complex argument with data and opinions is presented to convince a reader or listener. Toulmin, Rieke, and Janik (1984) defined argumentation as “the whole activity of making claims, challenging them, backing them up by producing reasons, criticizing those reasons, rebutting those criticisms, and so on” (p. 14). The importance of argumentation is found in both the Common Core State Standards (2010) and the Next Generation Science Standards (Achieve, 2013). These documents emphasize the need for students to know how to take a critical stance when confronted with an argument, evaluate the quality of what they read, see, or hear, and defend their claims with appropriate evidence and reasoning. It has been noted that argumentation skills actually help increase students’ achievement and content knowledge by requiring them to think deeply about the content, construct their own understanding of the content, and apply that understanding as they construct their arguments or critique those of others. The role of argumentation in supporting deeper learning is well documented (Pellegrino & Hilton, 2013).

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/UJRT.

In an earlier study that provided the foundation for the educational game developed in this study, middle school science teachers reported success in teaching the knowledge and skills related to scientific argumentation, with large effect sizes, but that the process was tedious and time-consuming (Bulgren & Ellis, 2012; Bulgren, Ellis, & Marquis, 2013). Both teachers and students had difficulty engaging in the process and putting it into practice. A recent Pew survey of Advanced Placement and National Writing Project teachers confirms this difficulty. While digital tools have supported the development of many writing skills, teachers still feel that additional resources are needed to support students' ability to develop strong arguments from facts (Purcell, Buchanan, & Friedrich, 2013).

The recognition of the time needed to teach and learn argumentation led to exploration of digital tools, with the goal of preparing students to more fluently engage in complex argumentation in science classes. We proposed the use of a game to support learning skills and knowledge related to scientific argumentation with an enjoyable and engaging resource. We hypothesized that a highly compelling game would provide students with sufficient practice in the skills and knowledge related to scientific argumentation to support learning the skill, and would engage them in a way that made the process of making decisions about science to be fun. The features that make Reason Racer compelling and fun for students as they address the skills of scientific argumentation are those that create a sense of flow during game play.

Literature Overview

Games Engage Learners

Teachers began utilizing the engagement potential of games long before the advent of computers and educational online games. Games, particularly competitive games, engage students in a fun and exciting way. Games create excitement and involvement with content that is difficult to replicate with other instructional strategies. Teachers use games as a pedagogical strategy, with other instruction, to engage students with content and reinforce learning. Online games have many of the same characteristics as traditional games and have the potential both to significantly increase the number of times a student engages with the content and to extend interaction with the content well beyond the school day. Research on the effect of technology-based games has consistently shown positive results regarding motivation, persistence, curiosity, attention, and attitude toward learning (Shin, Sutherland, Norris, & Soloway, 2012) and that students can have significantly higher cognitive gains when working with games when compared to traditional instruction (Vogel et al., 2006). This is because online games are specifically designed to "compel" students to play, drawing then into playing a game over and over (Garris, Ahlers, & Driskell, 2002; Koster, 2005; Lazzaro, 2004).

In her discussion of massively multiplayer online role-playing games (MMORPGs), Dickey (2007) described game features specifically designed to support intrinsic motivation. Her analysis focused on character design and the narrative environment within the game as players take on roles and participate in a story. A casual game format can be used when designing educational games that include features specifically selected to engage students in both content and process. Such games may be similar to online competitive games that are multiplayer but do not involve role-play. These design features include focused goals, ease of learning, simplicity of play, immediate feedback, quick rewards, and fast-paced yet forgiving game-play (Tams, 2006; Wallace & Robbins, 2006; Waugh, 2006). More recent studies continue to add support to this list of game features that engage youth. Hamlen (2013) confirms earlier work that shows youth are motivated by both the challenge and the thinking required by video games and the reward system. Evans, Norton, Chang, Deckard, and Balci (2013) provide additional support for the engagement potential of incentives. These features engage students to the extent that they experience a heightened emotional attachment to the activity, a sense of "flow."

The concept of "flow" is an accepted construct describing intense engagement in an activity (Csíkszentmihályi, 1975, 1997). Flow describes the experience of feeling totally involved in an activity. During a game, for example, the player achieves a state of total focus, complete immersion, and limited awareness of time. It is generally assumed that the experience involves intense involvement, concentration, and enjoyment. Authors of a series of studies, summarized by Hoffman and Novak (2009), suggest that flow is a complex process, influenced by many elements. Additional features demonstrated to elicit a sense of flow include interactivity (Choi, Kim, & Kim, 2000; Huang 2003, 2006; Skadberg & Kimmell 2004), ease of use and perceived usefulness (Agarwal & Karahanna, 2000; Hsu & Lu 2004; Sanchez-Franco 2006), and attractiveness, novelty, and playfulness (Agarwal & Karahanna, 2000; Choi, Kim, & Kim, 2000; Huang 2003, 2006; Skadberg & Kimmel, 2004). The motivation to play a game again is, from our perspective, a result of the emotional attachment created by the experience of flow. Features such as control, feedback, repetition, and set goals have been shown to influence the motivational aspect of games, that is, the drive to play the game again (Shin, Sutherland, Norris, & Soloway, 2012). The social context of the game has also been demonstrated to be important in creating the motivation to play (Choi & Kim, 2004; Hsu & Lu, 2004). Hsu and Lu (2004) hypothesize that the social influence is informational. Other game designers agree that competition is a significant contributor to the play experience (Koster, 2005). All these features parallel key characteristics identified as critical components of engaged learning (Jones, 1998; Jones, Valdez, Norakowski, & Rasmussen, 1994; Schlechty, 1997).

Von Ahn demonstrated the tremendous capacity of games, designed using these characteristics, to engage people in completing difficult tasks. Through games he demonstrated the power of human computation (Von Ahn, 2006; Von Ahn & Dabbish, 2004; Von Ahn, Kedia, & Blum, 2006). The Reason Racer game, inspired by Von Ahn, is based on the assumption that incorporating casual game characteristics into otherwise difficult learning tasks will increase the engagement and emotional attachment of players to successfully play the game and learn the skills associated with scientific argumentation.

Practice Leads to Fluency

Fluency, the speed and accuracy with which a student performs a skill, is directly related to the ability to apply that skill in a novel or more complex situation, which is identified as transfer of learning (Binder, 1996; Bower & Orgel, 1981; Hook, & Jones, 2002; Lindsley, 1990; Snyder, 1992; White, 1986). Regardless of the subject being taught, educators have long realized that the ability to retrieve information in an easy and fluid manner is one goal of learning (Bransford, Brown, & Cocking, 2000). Since the amount of information a student can attend to at any one time is limited, being able to easily retrieve certain elements of a task, or fluency, allows students to focus more attention on other components of the task. If learners struggle when retrieving information, they will be unable to focus attention on using their knowledge and skills for solving new problems in innovative ways.

Improving a skill to a functional level of fluency takes time and may require hundreds of instances of practice in retrieving a piece of information or executing a procedure (Pellegrino & Hilton, 2013). To achieve a level of performance in either a simple or complex cognitive operation that results in transfer of learning requires students to engage in sustained practice of a skill. Having multiple opportunities to practice a skill, within the context of instruction, is a key component in learning (Bransford, Brown, & Cocking, 2000).

Rate-based games, characterized by a high number of opportunities to practice a specific set of skills, provide a way to build fluency. Not only is there an opportunity to practice a skill multiple times, but the emotionally charged, competitive environment supports a level of engagement not typically found in most practice settings. The applied theoretical model for this research proposes that an understanding of and skills associated with components of argumentation will develop as a result of students playing the game many times in an engaging environment.

While practice and feedback are essential components in fostering for transfer of learning, Pellegrino and Hilton (2013) also discuss the role that motivation plays. In their review of interventions designed to produce transferable knowledge, they identify students' beliefs about their abilities as learners as an essential component of learning. Yaeger and Walton's (2011) review of social–psychological interventions suggests that feelings of self-efficacy as learners enhance deeper learning.

In another example, Blackwell, Trzesniewski, and Dweck (2007) showed that an intervention designed to change low-income minority seventh-grade students' self-descriptions as learners resulted in reversing a pattern of declining mathematics grades by the end of the school year. The results of this study suggest that sustained game play can have a positive effect on students' confidence and motivation to engage in science (their self-efficacy as learners).

Feedback Improves Performance

Pellegrino and Hilton (2013), in their summary on the education of transferable knowledge and skills in the 21st century, recommend the continuous use of formative feedback to inform not only the teacher but, more importantly, the student. Providing students with the results of their performance is directly related to their skill acquisition. Students acquire skills "much more rapidly if they receive feedback about the correctness of what they have done" (p. 80). Through feedback, learners can become aware of the relationship between their responses and learning outcomes, and can use strategies to achieve their academic goals (Zimmerman, 1990). Shute (2008) defined formative feedback as "information communicated to the learner that is intended to modify the learner's thinking or behavior for the purpose of improving learning" (p. 1). Feedback within games has been shown to produce better performance than games without feedback (Moreno & Mayer, 2005) and to engage the learner (Evans, Norton, Chang, Deckard, & Balci, 2013).

A characteristic of casual games, including Reason Racer, is the provision of immediate feedback with the purpose of providing players with strategies to improve their performance. The theoretical assumption is that the emotionally charged, competitive environment "compels" students to play a game over and over (Garris, Ahlers, & Driskell, 2002; Koster, 2005; Lazzaro, 2004), with the goal of improving. It is assumed that the game environment sets the stage for players to be very receptive to the feedback. This is particularly true if the feedback improves performance sufficiently to result in game-based rewards, or achievements, that allow players to be more competitive and to improve in the skills related to scientific argumentation.

Research Questions

The primary research questions addressed were: (a) Does playing Reason Racer within the context of middle school science instruction improve students' knowledge about scientific argumentation components and process? (b) Does it impact their feelings of confidence and motivation to engage in science when compared to students who did not play the game?

We were also interested in learning more about students' perceptions of their game experiences. Questions provided in a survey addressed (a) whether students who played the game felt that Reason Racer would help them learn science content and the practice of scientific argumentation and (b) whether the casual game format created a sense of engagement and flow.

Methods

Participants

The original sample size included eight implementation teachers and 906 middle school students from seven school districts in northeastern Kansas. These included rural, urban, and suburban districts. A subset of 670 students completed the pre- and posttests and 249 students completed 10 or more Reason Racer game-play sessions. Enrollment at the participating schools ranged from 135 to 888 students, with an average enrollment of 570 students. The population was 53.56% male and 46.44% female. Of this, an average of 31.8% of the students were identified as economically disadvantaged. Overall, a majority of the students were White (79.84%), with 8.43% Hispanic, 4.19% African American, and 7.57% listed as other. English language learners contributed 3.04% of the population, and 12.06% of the students were identified as having disabilities.

The teachers were the five (of the initial eight) participants who successfully implemented all components of the study. These teachers were volunteers to the project, and the comparison classes were selected by convenience. Each implementation teacher was requested to identify a comparison teacher and group of students in the same building, district, or region, resulting in comparison group of 460 middle school science students.

Treatment teachers. The treatment teachers were the five (of the initial eight) participants who successfully implemented each of the two components of the study. These components included (a) professional development and (b) providing students with routine opportunities to play the game. Professional development included a single day that addressed the instructional integration of scientific argumentation and Reason Racer, assistance in enrolling their students in the game, and access to online resources that supported scientific argumentation and the use of Reason Racer during science instruction (http://reasonracer.wikispaces.com). The second component required teachers to provided their students with the opportunity to play the game as a part of science instruction a minimum of 10 times over a six-week period.

The treatment teachers also participated in district-provided professional development activities prior to and during the school year, and implemented the district science curriculum as directed. The actual process of science instruction was unique to each of the seven districts.

These teachers were volunteers to the project, and the comparison classes were selected by convenience. Each treatment teacher was requested to identify a comparison teacher and group of students in the same building, district, or region, resulting in comparison group of 460 middle school science students.

Comparison teachers. The comparison teachers and their students differed from the treatment group by not having access to the single day of professional development and the online resources addressing the instructional integration of scientific argumentation and Reason Racer. As members of the comparison group, they participated in the pre- and posttest measures during the same week as the treatment teachers. For purposes of comparison, the teachers participated in district inservice activities and provided district-directed science instruction without any intervention from the project (i.e., business as usual). The curriculum and instruction provided by these teachers were not monitored.

Research Design

A mixed-methods quasi-experimental design was used with the student as the unit of study. A series of repeated measures analyses of variance were conducted.

The Reason Racer Game

The Reason Racer game contains four parts, each designed to engage players in skills and knowledge related to scientific argumentation. When setting up play for students, the teacher assigns the game by selecting from 40 different scenarios covering topics in earth and space, life, physical, and technology and engineering sciences. The different scenarios populate the content of the game's challenges. The teacher makes this selection by accessing the entire list of scenarios in the teacher portal, as seen in Figure 1. This space allows the teacher to rename a scenario and assign it to a class. This portal has additional functions such as uploading rosters, viewing students' performance data, and editing scenario assignments.

The first part of the game orients the players through a humorous 30-second video about the content of a particular scenario and elements of scientific argumentation. Each scenario is introduced by a different, engaging, video. The second part involves the players in a competitive, multiplayer rally-type race, alternating between challenges, or PitStops, and racing segments across a variety of racecourses. The PitStops require actions that are common to fast-paced games, such as matching, ranking, sorting, and discriminating, all within a rate-based game interface. Figure 2 shows six of the eight PitStops from one scenario as an example. This scenario claims that a new technology for biofuel production could utilize an enzyme found in a panda's digestive system to help convert plant matter to biofuel. As with this example, the PitStops contain the content of the game, requiring students to identify components or make decisions about the claim, evidence, reasoning, and challenges. Students attempt to move through each PitStop as quickly, and with as few errors, as possible. Even though all the players are engaged in the same scenario, each is presented with different items. This happens because more

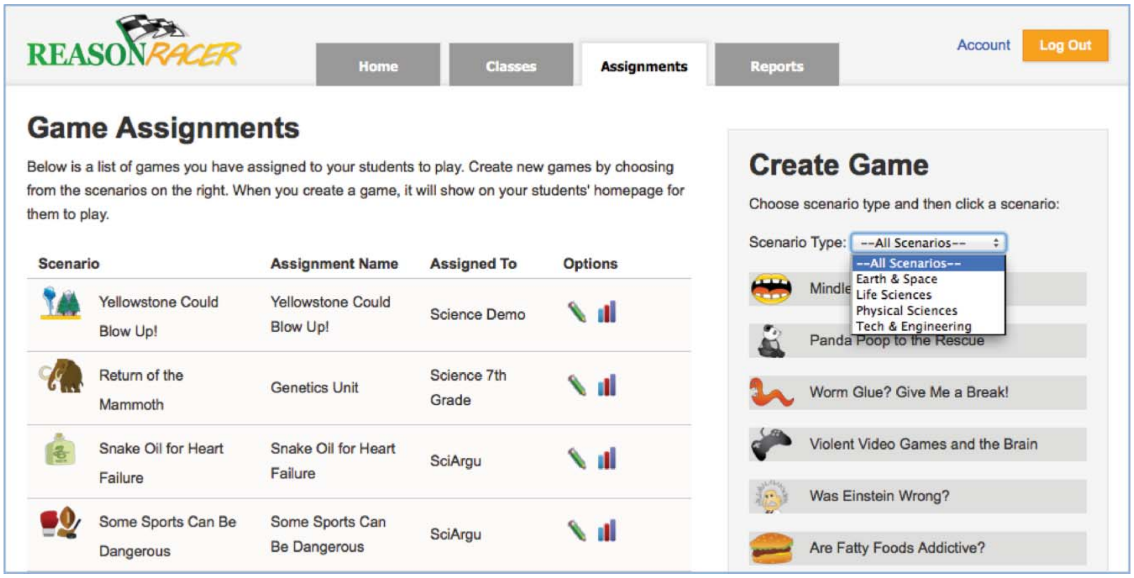

Figure 1. Teachers' Reason Racer portal, allowing for management of class assignments and review of students' performance.

items were generated per PitStop, per scenario, than one player would see in each game session. These items are randomly pulled from the database, with a unique distribution sequence to each player. This is similar to what might be seen in a randomized math game in which there is a standard problem type populated with many different items.

The competitive racing component, shown in Figure 3, is completed between PitStops. During this part of the game the students navigate various racing tracks with turns and obstacles as quickly as possible to move to the next PitStop. The speed and accuracy of a player's performance in the PitStop affects the speed with which his or her car can move through the next racing segment. Incorrect responses slow down the player's race car, which discourages guessing. The experience of the racing component occurring between the challenges has been demonstrated to result in students completing the PitStops more rapidly and more accurately (Ault, Craig-Hare, Bulgren, & Ellis, 2012).

Figure 2. A sample of PitStops that engage the players in actions that are common to fast-paced games such as matching, ranking, sorting, and discriminating, all within a rate-based game interface.

Figure 3. The staging area and sample of the racetracks that players navigate between PitStops.

The third part of the game involves decision making. Students read a brief article that reviews most of the content they just encountered while in the PitStops, an approximately 400-word essay. They move through the article at their own pace by using the space bar to scroll through the text. Their task, as shown in Figure 4, is to decide whether to accept, reject, or withhold their decision about the claim and write a justification for their decision. This comment is inserted into the final part of the game, the discourse portion (see Figure 5).

After making the decision and entering a statement supporting their decision, players race to the end of the game and receive their scores. This score also provides the player with achievements that allow him or her to be more competitive during the next play session.

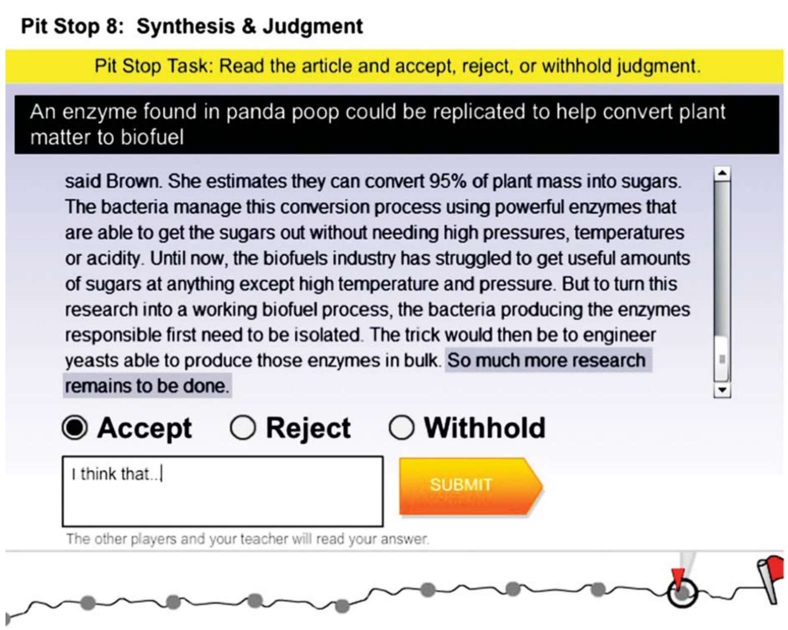

Figure 4. Decision portion of the game where students provide a rationale for their decision about the claim.

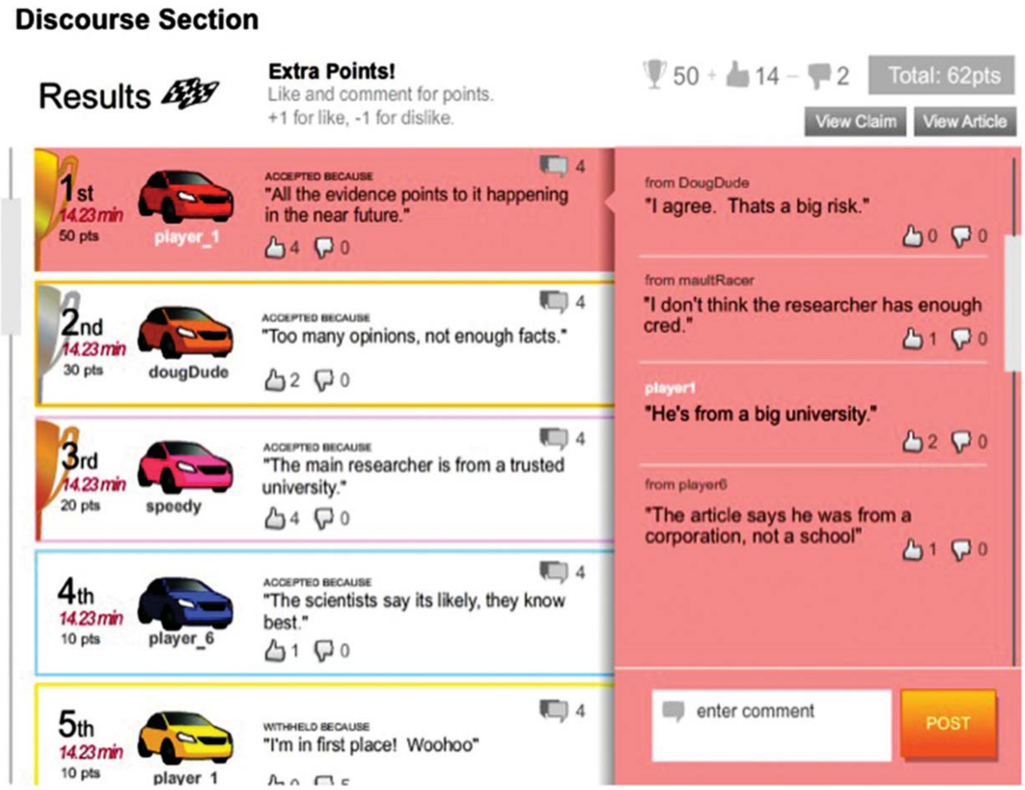

Figure 5. Discourse portion of the game where students interact with other players about decisions regarding the claim.

The fourth part of the game, Figure 4, involves a player engaging in discourse about his or her decision regarding a claim about a scenario with the other players who were involved in their race. During this discourse session, students are able to add to or subtract from other players' overall points in the game by scoring their comments. Students are encouraged to add points to players who make quality comments, and subtract from those who do not address the claim, evidence, or content of the article.

The multiplayer aspect of the game is evident in two different game actions. First, players are able to see the relative, real-time position of the other players in the game, represented as the small dots on the racetrack (see Figures 2 and 3). While players are not able to see other players' performance in a PitStop, they can see when players move out of a PitStop and are located in front of or behind their own car. Players actually compete against other players on the racetracks, in the racing portions of the game (see Figure 3). Each player sees his or her car on the racetrack with a slight highlight. Players are able to see as they pass each other and speed up their own car, or delay others cars through the use of their points. The multiplayer engine allows for 2 to 20 simultaneous players. The second game action that involves multiple players is in the discourse portion of the game. Here students can interact with others in the same game and add or remove points from others players' total scores.

Impact on Science Instruction

The teachers in the treatment condition arranged for their students to play Reason Racer during ongoing science class. Four of the teachers accessed a computer lab close to the science classroom; the other two teachers were located in a school with a 1:1 student-to-device initiative. In all cases the students accessed the game using a 1:1 ratio. Our request was that each teacher provide his or her students an opportunity to play the game a minimum of 10 times over a 6-week period. All the teachers differed in how they introduced and managed the game, varying in the number of game sessions per day, the number of scenarios assigned to the students, control over the make-up of the groups of students who played together, and the use of the teacher resources to review component of argumentation or how to play the game.

Data Collection

The materials used in this study included (a) the log files capturing the students' performance on each PitStop in the Reason Racer game, (b) a pre- and posttest addressing confidence and motivation related to knowing and learning science (see Appendix A and Appendix B), and (c) a survey addressing attitudes about the game. The game was accessed through the Internet, and individual student performance was captured in a database on the server. The pre- and posttest surveys and the survey about attitudes were also online, captured on a secure server, and provided during the same week to both the treatment and comparison groups.

Treatment group game performance. Game performance was collected as students participated in 10 separate game-play sessions, with the factor being the game-play sessions. The dependent variable was the number of errors made by students in each PitStop during each game-play session. Data were gathered in 5,897 game-play sessions from the six treatment classrooms. The analysis was conducted to determine whether the mean of the number of errors in each PitStop changed as students played more game sessions. The log files captured several metrics for each player during each game-play session, including duration in milliseconds for each component of the game; the number, frequency, and type of incorrect responses in each PitStop; and overall duration of the game.

Pretest and posttest measures. Two measures were taken comparing confidence and motivation between the treatment and comparison groups (see Appendix A). Science confidence was measured with a four-item Likert-type scale (reliability = .68). Each question began with "How confident are you that you . . ." and ended with specific science argumentation skills, such as explaining judgments about claims and having the knowledge and skills one needs to analyze claims. Answer options ranged from *not very confident* (1) to *very confident* (5). Science motivation was also measured with a four-item Likert-type scale (reliability = .78). Each question began with "How motivated are you to . . ." and ended with specific science argumentation activities, such as looking for information and evaluating evidence. Answer options ranged from *not very motivated* (1) to *very motivated* (5).

Treatment group's attitude toward the Reason Racer game. The treatment group members completed survey questions addressing their attitudes about the game. Attitude toward the game was measured with a 14-item Likert-type scale (reliability = .80). Each question asked the student to rate a statement from the first-person perspective, and began with "I found . . .," "I was . . .," "I felt . . .," and questions about the usefulness of the feedback. Answer options ranged from *strongly disagree* (1) to *strongly agree* (5). Four items on the scale were also analyzed for the presence of a sense of flow (reliability = .64) and included items such as "I found that time passed . . .," "I felt absorbed in . . .," and "I wanted to. . . ." Answer options ranged from *strongly disagree* (1) to *strongly agree* (5). The treatment group members also completed a survey asking about the game itself and whether they thought that playing the game helped them learn scientific argumentation. This survey included an eight-item Likert-type scale (reliability = .86). Each question began with the phrase "Playing the game helped me . . ." and conclusions like "comment on the views of others" or "engage in a discussion about claims." Answer options ranged from *strongly disagree* (1) *to strongly agree* (5).

Data Analysis

A series of repeated-measures analyses of variance was conducted. First, an overall analysis of variance, using all PitStops combined, with the 10 game-play sessions as the independent variable and the average number of wrong answers as the dependent variable was performed. Second, separate analyses of variance for each of four types of identification task questions (fact or opinion, qualifiers, reasoning type, and rebuttal, counterargument, or new question) were completed.

Results

Treatment Group Game Performance

Student game-play data were analyzed by reviewing error rates in seven of the eight PitStops across 10 game-play sessions. Because the eighth PitStop required students to enter their own decision to accept, reject, or withhold a claim and provide a justification, answers were not scored as incorrect or correct. These were, therefore, not included in this analysis. A repeated-measures analysis of variance (ANOVA) was conducted to examine the frequency and distribution of the student errors. The mean of errors was analyzed for each of the 10 game-play sessions for each type of task required in the PitStop. Overall, errors in game-play sessions decreased between sessions 1 and 10, except for the task requiring students to identify a statement as fact or opinion. In this instance, the mean of student errors slightly increased.

In addition the amount of time it took for students to play through a session decreased with practice, as would be anticipated. Students took an average of 11.5 minutes to complete the first game session and 6.7 minutes to complete the 10th session. The *SD* for the duration of both of these sessions was less than .01 seconds.

Performance in Each of the PitStops

Identify a statement as expressing a fact or opinion. The results for the repeated-measures ANOVA indicated a significant time effect, Wilks's $\Lambda = .93$, $F(9, 240) = 2.15$, $p < .05$, eta-squared $= .08$. The alternative univariate tests yielded the same F value, but corrected the degrees of freedom of the F as a function of the degree to which the data indicated that the sphericity assumption was violated. The p value was greater for the two alternative tests, but all three tests were significant at the traditional .05 level.

Determine the best claim statement for a short narrative. The results for the repeated-measures ANOVA indicated a significant time effect, Wilks's $\Lambda = .93$, $F(9, 240) = 1.99$, $p < .05$, eta-squared $= .07$. The alternative univariate tests yielded the same F value, but corrected the degrees of freedom of the F as a function of the degree to which the data indicated that the sphericity assumption was violated. The p value was slightly higher for the two alternative tests, but all three tests were significant at the traditional .05 level.

Identify the qualifiers in a statement. The results for the repeated-measures ANOVA indicated a significant time effect, Wilks's $\Lambda = .77$, $F(9, 240) = 8.02$, $p < .05$, eta-squared $= .23$. The alternative univariate tests yielded the same F value, but corrected the degrees of freedom of the F as a function of the degree to which the data indicated that the sphericity assumption was violated. The p value remained unchanged for the two alternative tests; all three tests were significant at the traditional .05 level.

Rank the evidence as it applies to a claim. The results for the repeated-measures ANOVA indicated a significant time effect, Wilks's $\Lambda = .93$, $F(9, 240) = 1.88$, $p < .05$, eta-squared $= .07$. The alternative univariate tests yielded the same F value, but corrected the degrees of freedom of the F as a function of the degree to which the data indicated that the sphericity assumption was violated. The p value was slightly higher for the two alternative tests, but all three tests were significant at the traditional .05 level.

Identify the type of reasoning. The results for the repeated-measures ANOVA indicated a significant time effect, Wilks's $\Lambda = .72$, $F(9, 240) = .72$, $p < .05$, eta-squared $= .28$. The alternative univariate tests yielded the same F value, but corrected the degrees of freedom of the F as a function of the degree to which the data indicated that the sphericity assumption was violated. The p value remained unchanged for the two alternative tests; all three tests were significant at the traditional .05 level.

Judge the strength of the reasoning as it applies to the claim. The results for the repeated-measures ANOVA indicated a significant time effect, Wilks's $\Lambda = .64$, $F(9, 240) = 15.03$, $p < .05$, eta-squared = .36. The alternative univariate tests yielded the same F value, but corrected the degrees of freedom of the F as a function of the degree to which the data indicated that the sphericity assumption was violated. The p value remained unchanged for the two alternative tests; all three tests were significant at the traditional .05 level.

Identify type of challenge (rebuttal, counterargument, new question) to the claim. The results for the repeated-measures ANOVA indicated a significant time effect, Wilks's $\Lambda = .80$, $F(9, 240) = 6.72$, $p < .05$, eta-squared = .20. The alternative univariate tests yielded the same F value, but corrected the degrees of freedom of the F as a function of the degree to which the data indicated that the sphericity assumption was violated. The p value remained unchanged for the two alternative tests; all three tests were significant at the traditional .05 level.

Follow-up pairwise comparisons for each PitStop indicated a significant linear effect with means decreasing over time and suggested that student error rates in each PitStop decreased through game-play sessions.

The eighth PitStop involved the player reading a summary article with a claim statement. The task is for the player to decide whether to accept, reject, or withhold the claim, and to provide a rationale statement. This PitStop does not provide the player with a score.

Summary of Errors by Task Type in Each PitStop: Identification or Judgment Task Type

Overall, the combined mean errors for all PitStops decreased across game-play sessions. This general trend is represented in Figure 6. There is some variability, however, in the distribution of errors across PitStops. Four of the PitStop tasks challenged students to *identify* elements of scientific argumentation, such as identifying a fact versus an opinion, qualifiers, types of reasoning, and the types of challenges to a claim. The mean of errors for performance in each of these four types of PitStops is reflected in Figure 7.

Three of the four *identification* tasks showed a decrease in errors across the 10 game-play sessions. A slight increase occurred in the PitStops that challenged students to identify a sentence as either a fact or opinion. The mean student error rate during the first game session in this PitStop was 1.36, and 1.37 during the 10th game session. The mean error rates in the other three identification tasks decreased in across game-play sessions. The task asking students to identify types of challenges to a claim had an error rate of 2.17 during the first game-play session and 1.62 during the 10th game-play session. The tasks asking students to identify qualifiers and the types of reasoning appeared to be more difficult for students. The first game-play session reflected error rates of 2.98 and 4.52, dropping to .53 and 3.15, respectively, as students mastered these elements throughout game-play sessions.

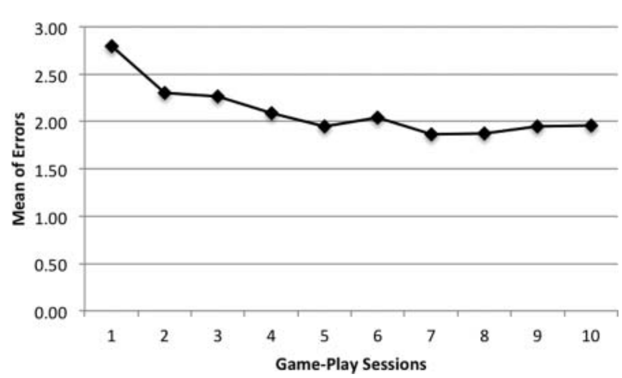

Figure 6. Mean of errors for all PitStops across game-play sessions.

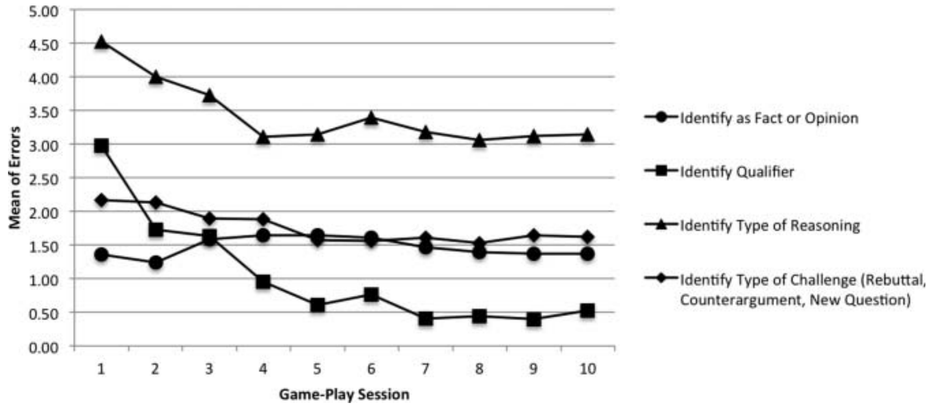

Figure 7. Mean of errors in identify tasks.

Higher order thinking tasks, such as those requiring students to make judgments about information, also reflected a decrease in error rates as students played the game over multiple sessions (Figure 8). Similar to the identification tasks, all the *judgment* tasks indicated a decrease in mean error rates from the first to the 10th game-play sessions. The mean error rate for the task challenging students to determine the best claim statement changed from a value of 2.01 in the first game-play session to 1.69 during the 10th game session. The task challenging students to rank evidence as it applied to a claim went from a mean error rate of 3.06 during the first game session to 2.51 during the 10th session. Similarly, students decreased their mean error rates in judging the strength of reasoning as it applied to a claim, beginning with their first game session having an error rate of 3.46 and ending with their 10th session at 2.84 mean errors.

Pretest and Posttest Measures

Measure of confidence regarding science. A science confidence test was completed by 739 middle grade science students in the implementation group who took part in the Reason Racer evidence game intervention, and 240 middle grade science students in the comparison group. The scale was administered twice, at the beginning of the semester before Reason Racer was introduced to the class, and at the end of the semester after use of the game was over. A repeated-measures analysis of variance was conducted with either the treatment or comparison group as the between-subjects factor, time between measurement occasions as the within-subjects factor, and total mean score on the confidence scale as the dependent variable. Results were significant, with students in the treatment group gaining confidence to a greater degree than students in the comparison group, $F(1,977) = 14.75$, $p \leq .001$, with a small effect size, partial eta squared = .015. The descriptive statistics are reported in Table 1, and this relationship is reflected graphically in Figure 9.

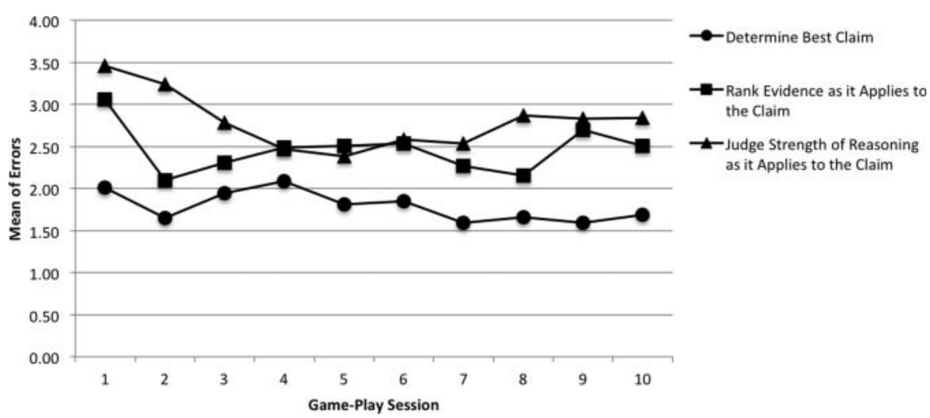

Figure 8. Mean of errors in judgment tasks.

Table 1. Comparison of Science Confidence Scores

Group	*n*	Pretest		Posttest	
		Mean	*SD*	Mean	*SD*
Implementation	739	3.41	.75	3.68	.66
Comparison	240	3.34	.69	3.38	.66

Note. Measured on a 1 to 5 scale with higher scores indicating greater confidence.

Measure of motivation regarding science. The science motivation scale also was administered twice, at the same time as the confidence scale, completed by the same 739 treatment and 240 comparison middle school science students. A repeated-measures analysis of variance was conducted with the treatment or comparison group as the between-subjects factor, time between measurement occasions as the within-subjects factor, and total mean score on the motivation scale as the dependent variable. Results were significant, with students in the treatment group showing increased motivation to a greater degree than students in the comparison group, $F(1{,}977) = 11.77$, $p = .001$, with a small effect size, partial eta squared = .012. The descriptive statistics are reported in Table 2, and this relationship is reflected graphically in Figure 10.

Treatment Group's Attitude Toward Reason Racer

The treatment group, those students who used the Reason Racer game, completed a survey on their attitude toward the game (see Table 3) and their experience of a sense of flow while playing (see Table 4). The mean total for questions addressing whether they believed the game helped them engage in the components of scientific argumentation was 3.54 (on a 5-point scale). The mean for items addressing the players' attitude toward the game, however, was 3.74. The mean for items addressing a sense of flow (items 4, 7, and 14) was 3.53. (Note: Items 4, 5, 7, 8, 11 and 12 were reversed before totaling and before reliability analysis.)

Discussion

Key Findings

Within the context of science instruction, middle school students who played Reason Racer multiple times improved in their performance in each of the PitStops, all of which addressed a component of scientific argumentation. In other words, performance improved with each successive game-play session. That is, the more the students played the game, the more their performance improved in each PitStop, as reflected by the continued reduction in errors.

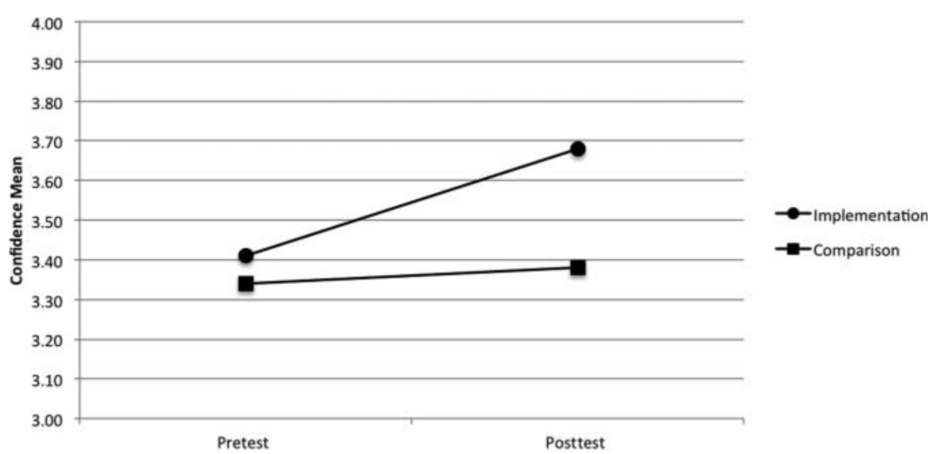

Figure 9. Comparison of pre- and posttest mean scores of implementation and comparison groups' confidence in their ability to engage in science. Measured on a 1 to 5 scale with higher scores indicating greater motivation.

Table 2. Comparison of Science Motivation Scores

		Pretest		Posttest	
Group	*n*	Mean	*SD*	Mean	*SD*
Implementation	739	3.20	.87	3.35	.84
Comparison	240	3.02	.78	2.94	.83

Two additional findings addressed students' responses to playing the game. During pre- and posttest comparisons, students who played Reason Racer as a part of middle school instruction reflected a mild but significant increase in their confidence and motivation to engage in science compared to students who did not play the game. Additionally, the students who played Reason Racer reported a positive attitude toward the game. They also reported feelings while playing the game that are associated with a sense of flow.

This study demonstrated that a large number of middle school students across urban, suburban, and rural school districts improved performance in all components of a game, Reason Racer, addressing the skill of scientific argumentation as outlined by Toulmin (Toulmin, 2003; Toulmin, Rieke, & Janik, 1984). These findings support two of the three original assumptions about the effectiveness of a game incorporating a specific set of features. The findings contribute to the evidence that games incorporating a casual-game format including focused goals, simple play, quick rewards, and fast-paced or timed game-play (Tams, 2006; Wallace & Robbins, 2006; Waugh, 2006) and competition (Koster, 2005) can engage students and create a sense of emotional attachment, or flow, when played. Students who played the game reported feeling that they experienced a sense of flow such as rapid passage of time and feelings of being absorbed by the game (3.53 on a 5-point scale).

Students' positive feelings about the game were also indicated by their reports of confidence and motivation in doing science, compared to students who did not play the game as a part of science instruction. Motivation, identified by students' beliefs about their abilities as learners, is considered to have a significant role in learning (Pellegrino & Hilton, 2013). Yaeger and Walton's (2011) review of social–psychological interventions suggests that feelings of self-efficacy as learners enhance deeper learning. These findings are consistent with other findings on the effects of technology games in improving motivation, persistence, curiosity, attention, and attitude toward learning (Shin, Sutherland, Norris, & Soloway, 2012). Dickey (2007) concluded that the narrative environment within a game, as players take on roles and participate in a story, is a critical component for motivation to continue to engage players in the game. The findings in this study suggest that other game features can produce the motivation to continue play, and engage players in the content. The

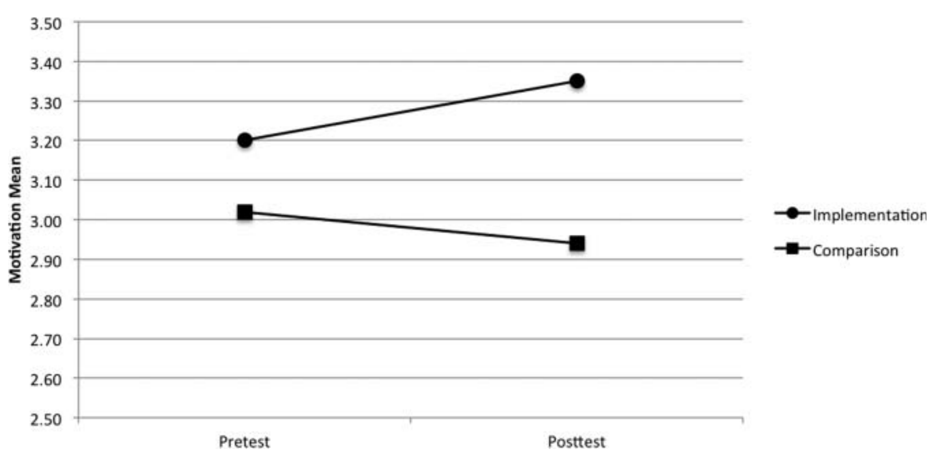

Figure 10. Comparison of pre- and posttest mean scores of implementation and comparison groups' motivation to engage in science. Measured on a 1 to 5 scale with higher scores indicating greater motivation.

Table 3. Attitude Survey

Score: 1 equals *strongly disagree* and 5 equals *strongly agree*.

Thinking about the Reason Racer game, please indicate the level to which you agree with the following statements:

Item	Mean	*SD*
1. I knew what I had to do to complete the game.	4.18	.89
2. I wanted to complete the game.	4.20	.91
3. I found it easy to get started.	4.00	.98
4. I found the game boring.	2.23	1.11
5. I found the game frustrating.	2.53	1.18
6. I was interested in exploring the options in the game.	3.45	.97
7. I felt that time passed slowly.	2.53	1.07
8. It wasn't clear what I could and couldn't do.	2.29	1.04
9. I felt that I could achieve the goal of the game.	4.03	.92
10. What I could learn from the game was clear.	3.87	.96
11. I could not tell what effect my actions had.	2.64	1.02
12. I did not care how the game ended.	2.60	1.17
13. Feedback I was given was useful.	3.40	1.01
14. I felt absorbed in the game.	3.38	1.10

$N = 734$ students in the implementation group
Mean total attitude toward game = 3.74 (reliability = .80)
Mean total *flow* (items 4, 7 and 14) = 3.53 (reliability = .64)

Note. Items 4, 5, 7, 8, 11, and 12 were reversed before totaling and before reliability analysis.

results also suggest that playing a game with casual-game features has a positive effect on students' confidence and motivation to engage in the content of the game; the content being science in this case. The use of these relatively inexpensive production strategies in a game provides options for creating the motivation needed to engage learners.

These results also contribute to the evidence that educational games can provide increasing amounts of practice in an engaging environment and have the potential to result in improved performance. The importance of practice in learning a skill, to the point of transfer, is well founded. To achieve a level of performance of either a simple or complex cognitive operation that results in transfer of learning requires students to engage in sustained practice of a skill (Bransford, Brown, & Cocking, 2000). The ability to develop a skill to a functional level of fluency to allow for the transfer takes time, and may require hundreds of instances of practice in retrieving a piece of information or executing a procedure (Pellegrino & Hilton, 2013). The Reason Racer study demonstrated a strategy in which students engaged in more than 400 decisions based on Toulmin's framework related to scientific argumentation across 10 game-play sessions. Results showed that students consistently improved performance across the first five game-play sessions, with overall improvements across

Table 4. Belief Survey

Item	Mean	*SD*
Playing the game helped me . . .		
1 . . .understand how to think about scientific argumentation in class.	3.55	0.98
2 . . .think about scientific argumentation outside of class, such as from newspapers, the Internet, or television news.	3.36	1.01
3 . . .create rebuttals, counterarguments, or new questions about a claim.	3.35	0.96
4 . . .come to a conclusion about a claim.	3.67	0.90
5 . . .comment on others' evaluations of a claim.	3.66	0.97
6 . . .engage in a discussion about claims and argumentation.	3.49	0.99
7 . . .be more comfortable engaging in discussions with my peers.	3.56	1.02
8 . . .give feedback to others.	3.69	1.03

Mean total belief that game helped = 3.54 (reliability = .86) the 10 game-play sessions. This improvement occurred both for the PitStops that involved identification tasks and those that engaged the students in higher-order thinking. These higher order tasks required the students to make judgments about the best claim statement, to rank evidence, and to determine the strength of the reasoning, all as applied to a claim. Most importantly, these findings suggest that game features typically associated with drill and practice games can be used to engage students and improve their performance in higher order thinking skills.

Limitations of the Study

We believe there are two major limitations to this study. First, the treatment teachers were volunteers and their comparison classroom teachers were selected based on convenience. Even though the sample size of students is large and represents urban, suburban, and rural school districts, the generalizability of the findings is limited to middle school students.

Second, a portion of teachers, and therefore students, did not complete the requirements of participation for this study. This included participation in professional development activities, integrating Reason Racer into ongoing instruction with a minimum of 10 game-play sessions in one semester, and completing both the pre- and posttest. The study started with eight teachers and 906 students. Five of the eight teachers engaged their students in 10 or more game-play sessions. Individual follow-up discussions with each of the teachers revealed that those teachers who did provide 10 or more practice sessions also reported integrating discussions about the components of argumentation into their science instruction. The three teachers who did not provide their students with the opportunity to play the game the required 10 times reported that they did not integrate argumentation into any part of their science instruction. These three teachers also did not fully participate in the professional development. The reasons they did not participate varied, but based on their reports were the result of personal issues such as conflicting schedules, no access to technology, illness, and so on. The difference in the number of students who completed the pre-and posttests, 670, and the 249 who completed a minimum of 10 Reason Racer game-play sessions can be accounted for by the three teachers who did not complete the requirements of the study, in addition to students who were absent from class for a variety of reasons. Teacher interviews revealed that insufficient instructional time and special events may not have allowed for all teachers to give all students the opportunity to play 10 sessions.

Implications for Practice

Inquiry practice, such as argumentation, is a difficult process to include in science instruction due to students' misconceptions about the nature and creation of knowledge. Yet increasingly, science standards and instruction challenge students to "reason scientifically," as specified in *America's Lab Report: Investigations in High School Science* (Singer, Hilton, & Schweingruber, 2006), *How Students Learn Science in the Classroom* (NRC, 2005), *Taking Science to School* (Duschl, Schweingruber, & Shouse, 2007), *Framework for K-12 Science Education* (Quinn, Schweingruber, Keller, 2011), and the *Next Generation Science Standards* (Achieve, 2012). Students, in their discussions about science, do not differentiate their opinions from other types of evidence, nor do traditional classroom norms support discussion about types of evidence (Kuhn & Reiser, 2005). This is because the process of engaging students in scientific argumentation is complex and challenging (Kuhn & Reiser, 2005; McNeill, Lizotte, Krajcik, & Marx, 2006). A resource that supports students in engaging in and thinking about the process of scientific argumentation will assist teachers in addressing the development of this skill within middle school science instruction.

The Reason Racer game is specifically designed to provide students with practice in identifying and making decisions about the evidence and reasoning supporting a scientific claim within a highly engaging game. Integrating Reason Racer into middle school science instruction requires minimal class time, and teachers may use the game as a preview, review, or supplemental activity to their traditional instruction. An entire class can participate in small-group games lasting less than 10 minutes. Even though the game will support 20 simultaneous players, teachers report that smaller groups of four to six players result in more excitement and collaboration within the class. Teachers can select game content that is aligned with their curriculum by selecting from more than 40 scenarios in all areas of middle school science. Teachers also have access to a comprehensive reporting system for individual and overall class performance on each PitStop. This system provides information on the most common misconceptions held by the class, as well as on individual student achievement.

Recommendations for Future Research

This research contributes to the understanding of how educational games using a targeted-game format might engage, motivate, and enhance the learning of higher order thinking in science education (e.g. scientific argumentation). Further research should address the transfer of game performance in the components of scientific argumentation to other content areas. As is reflected in the Common Core State Standards (2010), students need to know how to take a more critical stance when confronted with an argument, evaluate the quality of what they read, see or hear, and defend their claims with appropriate evidence and reasoning. Research should investigate the possible transfer of the scientific argumentation skills developed as a result of participation in Reason Racer to discussions in other classes, such as mathematics, social studies, or language arts.

The effect of various components of Reason Racer should be examined further. One area of further study should address whether there is an effect as a result of the number of different scenarios used by the teachers. In this study, teachers had total control over selecting and assigning scenarios for their students. Some of the teachers limited the content to two or three scenarios, while other teachers assigned 10 or more scenarios and allowed students to select the content they would address in the game. The scenarios represented different content as well as different Lexile Framework for Reading scores. An analysis should determine whether the teachers' pattern of assignment of scenarios affected students' performance. Another unknown is the degree to which students used the in-game prompts or feedback to improve their performance. Given the tremendous amount of data that is generated during game play, how could these data be represented to the students in order to increase their strategic use of the data, and therefore their success in the game? Additional research might address the role that a casual-game format has in improving students' attitudes toward their understanding of and ability to develop skills in science. More specifically, does the positive emotional attachment to the game influence students' attitudes towards its content? An additional question might address the types of feedback that could be provided to students to enhance their game play. Finally, further research should address whether similar types of games, ones that are specifically designed to heighten emotional attachment, can be an integral part of the instruction of complex and difficult skills. Recognizing that many complex skills are challenging and difficult, engaging students in these tasks takes advantage of what Von Ahn has observed: "Some tasks are inherently unenjoyable—until you make them a game" (Thompson, 2007, p. 4).

Received: 8/30/13
Initial decision: 6/25/14
Revised manuscript accepted: 8/8/14

Acknowledgments. The authors thank the students and teachers from Argentine Middle School in Kansas City, Kansas who assisted with the game and resource development as participant designers. They also thank the middle school science teachers and students who participated in research activities from the Auburn-Washburn, Blue Valley, Gardner Edgerton, Kansas City, North Jackson, Shawnee Heights, and Shawnee Mission School Districts in Kansas

Declaration of Conflicting Interests. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding. This research was supported by the National Science Foundation under grant 1019842, The Evidence Games: Collaborative Games Engaging Middle School Students in the Evaluation of

Scientific Evidence. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Author Notes

Marilyn Ault is an associate research scientist and director of ALTEC at the University of Kansas Center for Research on Learning. Her research interests focus on the instructional use of technology, including the design and use of educational games, to affect teaching and learning. Please address correspondence regarding this article to Marilyn Ault, ALTEC, University of Kansas, 1122 West Campus Road, Rm. 748, Lawrence, KS 66045, USA. E-mail: mault@ku.edu.

Jana Craig-Hare is an assistant research professor and associate director of ALTEC at the University of Kansas Center for Research on Learning. Her research interests focus on the instructional use of technology, school-wide interventions, and program evaluation.

Bruce Frey is an associate professor in the Department of Psychology and Research in Education at the University of Kansas. His research interests focus on classroom assessment, instrument development, and program evaluation methodology.

James D. Ellis is an associate professor in the Department of Curriculum and Teaching at the University of Kansas. His research interests focus on reform in science education, science teacher education, and use of technology to enhance teaching and learning in science.

Janis Bulgren is a research professor in the University of Kansas Center for Research on Learning. Her research interests focus on the development of and research on instructional procedures across the content areas that help students engage in higher order reasoning.

References

Achieve, Inc. (2013). *Next Generation Science standards*. Retrieved fromhttp://www.nextgenscience.org

Atkins, L. J. (2008, October). The roles of evidence in scientific argument. *AIP Conference Proceedings*, *1064*, 63.

Ault, M., Craig Hare, J., Bulgren, J., & Ellis, J. (2012). Analysis of a game feature designed to heighten engagement and learning. In P. Resta (Ed.), Proceedings of Society for Information Technology & Teacher Education International Conference 2012 (2462–2466). Chesapeake, VA: AACE. Retrieved fromhttp://www.editlib.org/p/39952

Binder, C. V. (1996). Behavioral fluency: Evolution of a new paradigm. *Behavior Analyst*, *19*, 163–197.

Blackwell, L. A., Trzesniewski, K. H., & Dweck, C. S. (2007). Theories of intelligence and achievement across the junior high school transition: A longitudinal study and an intervention. *Child Development*, *78*, 246–263.

Bower, B., & Orgel, R. (1981). To err is divine. *Journal of Precision Teaching*, *2*(l), 3–12.

Bransford, J., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school.* Washington, DC: National Research Council Committee on Learning Research and Educational Practice, National Research Council.

Bulgren, J. A., & Ellis, J. D. (2012). Argumentation and evaluation intervention in science classes: Teaching and learning with Toulmin. In M. S. Kline (Ed.), *Perspectives on scientific argumentation: Theory, practice, and research* (pp. 135–154). New York, NY: Springer.

Bulgren, J. A., Ellis, J. D., & Marquis, J. (2013). The use and effectiveness of an argumentation and evaluation intervention in science classes. *Journal of Science Education Technology.* doi:10.1007/s10956-013-9452-x

Choi, D., & Kim, J. (2004). Why people continue to play online games: In search of critical design factors to increase customer loyalty to online contents. *Cyberpsychology and Behavior*, *7*(1), 11–24.

Choi, D., Kim, H., & Kim, J. (2000). A cognitive and emotional strategy for computer game design. *Journal of MIS Research*, *10*, 165–187.

Common Core State Standards. (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Csíszentmiháyi, M. (1975). *Beyond boredom and anxiety.* San Francisco, CA: Jossey-Bass.

Csíszentmiháyi, M. (1997). *Finding flow: The psychology of engagement with everyday life.* New York, NY: Basic Books.

Dickey, M. D. (2007). Game design and learning: A conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation. *Educational Technology Research and Development*, *55*(3), 253–273.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academies Press.

Evans, M., Norton, A., Chang, M., Deckard, K., & Balci, O. (2013). Youth and video games: Exploring the effects of learning, achievement, and engagement. *Zeitschrift für Psychologie*, *221*(2), 98–106.

Garris, R., Ahlers, R., & Driskell, J. (2002). Games, motivation and learning: A research and practice model. *Simulation and Gaming*, *33*, 441–467.

Hamlen, K. R. (2013). Understanding children's choices and cognition in video game play: A synthesis of three studies. *Zeitschrift für Psychologie*, *221*(2), 107.

Hoffman, D. L., & Novak, T. P. (2009). Flow online: Lessons learned and future prospects. *Journal of Interactive Marketing*, *23*(1), 23–34.

Hook, P. E., & Jones, S. D. (2002). The importance of automaticity and fluency for efficient reading comprehension. *Perspectives on Language and Literacy*, *28*(1), 9–14.

Hsu, C. L., & Lu, H. P. (2004). Why do people play on-line games? An extended TAM with social influences and flow experience. *Information and Management*, *41*, 853–868.

Huang, M. H. (2003). Designing website attributes to induce experiential encounters. *Computers in Human Behavior*, *19*, 425–442.

Huang, M. H. (2006). Flow, enduring and situational involvement in the Web environment: A tripartite second-order examination. *Psychology and Marketing*, *23*, 383–411.

Jones, B., Valdez, G., Norakowski, J., & Rasmussen, C. (1994). *Designing learning and technology for educational reform*. North Central Regional Educational Laboratory. Retrieved fromhttp:www.ncrtec.org/capacity/profile/profwww.htm

Jones, M. G. (1998). Creating engagement in computer-based learning environments. *ITForum*. Retrieved fromhttp://itech1.coe.uga.edu/itforum/paper30/paper30.html

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

Kuhn, L., & Reiser, B. (2005, April). Students constructing and defending evidence-based scientific explanations. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Dallas, TX.

Lazzaro, N. (2004). *Why we play games: Four keys to more emotion without story.* Retrieved fromhttp://www.xeodesign.com/xeodesign_whyweplaygames.pdf

Lindsley, O. R. (1990). Precision teaching: By teachers for children. *Teaching Exceptional Children*, *22*(3), 10–15.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R.W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *Journal of the Learning Sciences*, *15*(2), 153–191.

Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, *97*(1), 117.

National Research Council. (2005). *How students learn: Science in the classroom*. Washington, DC: National Academies Press.

Pellegrino, J. W., & Hilton, M. L. (2013). *Education for life and work: Developing transferable knowledge and skills in the 21st century.* Washington, DC: National Academies Press.

Purcell, K., Buchanan, J., & Friedrich, L. (2013). *The impact of digital tools on student writing and how writing is taught in schools*. The Pew Research Center, National Writing Project. Retrieved fromhttp://pewinternet.org/Reports/2013/Teachers-technology-and-writing

Quinn, H., Schweingruber, H., & Keller, T. (Eds.). (2011). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

Sanchez-Franco, M. J. (2006). Exploring the influence of gender on web usage via partial least squares. *Behavior and Information Technology*, *25*(1), 19–36.

Schlechty, P. C. (1997). *Inventing better schools: An action plan for educational reform*. San Francisco, CA: Jossey-Bass.

Shin, N., Sutherland, L. M., Norris, C. A., & Soloway, E. (2012). Effects of game technology on elementary student learning in mathematics. *British Journal of Educational Technology*, *43*(4), 540–560.

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

Singer, S. R., Hilton, M. L., & Schweingruber, H. A. (Eds.). (2006). *America's lab report: Investigations in high school science*. Washington, DC: National Academies Press.

Skadberg, Y. X., & Kimmel, J. R. (2004). Visitors' flow experience while browsing a Web site: Its measurement, contributing factors and consequences. *Computers in Human Behavior*, *20*, 403–22.

Snyder, G. (1992). Training to fluency: A real return on investment. *Performance Management Magazine*, *10*, 16–22.

Tams, J. (2006). Online casual games Q&A. *Minna Magazine*, 2–5. Retrieved fromhttp://mag.casualconnect.org/MinnaMagazine_Summer2006.pdf

Thompson, C. (2007). For certain tasks, the cortex still beats the CPU. *Wired Magazine*, *15.07*, 4. Retrieved from http://archive.wired.com/techbiz/it/magazine/15–07/ff_humancomp?currentPage=all

Toulmin, S. (2003). *The uses of argument*. Cambridge, UK: Cambridge University Press.

Toulmin, S., Rieke, R., & Janik, A. (1984). *An introduction to reasoning*. Upper Saddle River, NJ: Prentice Hall.

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.

Von Ahn, L. (2006). Games with a purpose. *Computer*, *29*(6), 92–94.

Von Ahn, L., & Dabbish, L. (2004). Labeling images with a computer game. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 319–326). New York, NY: ACM Press.

Von Ahn, L., Kedia, M., & Blum, M. (2006). Verbosity: A game for collecting commonsense facts. *Proceedings of the SIG-CHI Conference on Human Factors in Computing Systems* (pp. 75–78). New York, NY: ACM Press.

Wallace, M., & Robbins, B. (Eds.). (2006). *IGDA 2006 casual games white paper*. Retrieved fromhttp://www.igda.org/casual/IGDA_CasualGames_Whitepaper_2006.pdf

Waugh, E. (2006). *GDC: Casual Games Summit 2006: An introduction to casual games*. Retrieved fromhttp://www.gamasutra.com/features/20060322/waugh_01.shtml

White, O. R. (1986). Precision teaching: Precision learning. *Exceptional Children*, *52*(6), 522–534.

Yaeger, D. S., & Walton, G. M. (2011). Social–psychological interventions in education: They're not magic. *Review of Educational Research*, *81*, 267–301.

Zimmerman, B. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, *25*, 3–17.

Appendix A: Confidence Survey

Please indicate how confident you are that you can think about scientific claims, evidence, and reasoning. Scientific claims are statements made about an important issue and present evidence and reasoning why others should accept the claim. Examples may range from a conclusion from a laboratory experiment on motion of a falling object to an article claiming that we should eat genetically-engineered food. Answer each of the following questions by selecting the button that best describes your response to scientific claims.
How confident are you that you . . .

	Not Very Confident	Not Confident	Neither Unconfident nor Confident	Confident	Very Confident
have the knowledge and skills you need to analyze and make strong scientific claims?	1	2	3	4	5
have a process or know a series of steps to go through as you analyze or make a claim dealing with a science issue?	1	2	3	4	5
are correct in your decisions about whether to accept or reject scientific claims you hear or read about?	1	2	3	4	5
Can explain your judgments about scientific claims to others?	1	2	3	4	5

Appendix B: Motivation Survey

Please indicate how motivated you are that you can think about scientific claims, evidence, and reasoning.
How motivated are you to . . .

	Not Very Motivated	Not Motivated	Neither Unmotivated nor Motivated	Motivated	Very Motivated
look for articles and information about scientific claims?	1	2	3	4	5
engage in discussions about scientific claims?	1	2	3	4	5
evaluate the evidence and reasoning made in support of claims?	1	2	3	4	5
explain your judgments about whether to accept or reject a claim to others?	1	2	3	4	5