CRESST REPORT 804

THE USE OF NARRATIVE: GENDER DIFFERENCES AND IMPLICATIONS FOR MOTIVATION AND LEARNING IN A MATH GAME

AUGUST, 2011

Sarah Joy Bittick

Gregory K.W.K. Chung

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

UCLA | Graduate School of Education & Information Studies

The Use of Narrative: Gender Differences and Implications for Motivation and Learning in a Math Game

CRESST Report 804

Sarah Joy Bittick and Gregory K. W. K. Chung
CRESST/University of California, Los Angeles

August, 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: Bittick, S.J., & Chung, G.K.W.K.. (2011). *The use of narrative: Gender differences and implications for motivation and learning in a math game*. (CRESST Report 804). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

TABLE OF CONTENTS

THE USE OF NARRATIVE: GENDER DIFFERENCES AND IMPLICATIONS FOR MOTIVATION AND LEARNING IN A MATH GAME

Sarah Joy Bittick and Gregory K. W. K. Chung
CRESST/University of California, Los Angeles

Abstract

The use of a narrative in educational contexts has been found to increase learners' experience of flow or absorption in a task. This increased experience of flow can in turn result in increased retention and learning outcomes. However, narrative can also be polarizing particularly in the male-dominated realm of video game play due to gender stereotyping or underrepresentation of females in games. Considering studies on narrative, flow, and gender differences in video games, it is expected that masculine and feminine narratives used in an educational game would result in differences in learning for the two genders. This study implemented two narratives into a rational numbers math video game to compare engagement and learning outcomes between genders and in comparison to a non-narrative version of the game. One hundred seventy-seven students enrolled in a remedial math course, pre-algebra, or Algebra 1 participated in the main study; this included 85 males, 80 females, and 12 students who did not report their gender. Results indicate that narrative did increase student experience of flow and positive perceptions of the game, especially when males were matched with the masculine version of the game. Increased learning outcomes took place only when students were placed in the masculine narrative and when males were matched to the masculine narrative.

Introduction

Background

Narrative can be a potent instructional strategy to facilitate how students encode, recall, and transfer learned material. Many gaming environments use narrative to communicate to the player the context of the game situation. For example, narrative is often used to provide the player with a sense of who their character is, where they are situated, what challenges await them and why, and what resources (human or otherwise) are available to overcome those challenges. From an entertainment perspective, the use of narrative can serve to enhance player motivation, increase player engagement, and affect how a player values completing various tasks or missions within a game.

Studies involving narrative demonstrate the overwhelming impact the presence of a narrative can have on the engagement a person experiences while participating in the activity. Experience of flow, transportation, presence, or absorption are terms used by researchers to describe the experience of being “lost” in a activity (Busselle & Bilandzic, 2008; Csikszentmihalyi, 1988; Green, Brock, & Kaufman, 2004; Park, Lee, Jin, & Kang, 2010). These experiences are measured with a variety of affective responses or outcomes. For example, in a study of a computer-based business course, 40% of learners reported the integrated details about the story and course content when asked what was most interesting about the course (Bielenberg & Carpenter-Smith, 1997). Some studies have specifically considered the role of narrative in video games, often using “presence” as a measure of student engagement. Park and his colleagues found that students given a pre-game story in the form of a video reported increased positive views of the video game and higher feelings of presence while playing (Park et al., 2010). In a study comparing story-based and non-story-based versions of a first person shooter game, results indicated that students receiving the narrative versions reported higher identification with the character, feelings of presence, and more positive emotional responses to the game (Schneider, Lang, Shin, & Bradley, 2004).

It is necessary to combine educational theory and game design so that learning games will be both effective and engaging (Kiili, 2005). This might be achieved through the use of narrative in educational gaming as narrative has been shown in the studies mentioned above to increase engagement. There seems to be some inconsistency in the literature as to whether engagement will, in fact, increase student learning. One study of a hypermedia learning environment did not observe increased learning when students experienced higher levels of flow (Konradt & Sulz, 2001). However, other studies including one involving game-based learning found that learning was positively impacted by increased flow experience (Skadberg & Kimmel, 2004; Webster, Trevino, & Ryan, 1993). Additionally, there is evidence suggesting that the use of narrative in educational games can improve student performance, on a posttest for example, and increase their self-perceived understanding of the game (Koenig, 2008). A goal of this study is to assess the learning differences that may arise when students are given a narrative in a math video game.

Unfortunately, not just any story can be inserted into a learning game; its themes and representations of characters must be taken into consideration. Infusing an explicit narrative into an educational game does not guarantee engagement or value on the part of the player as an explicit narrative can be polarizing to some players (Koenig, 2007). This may be partially due to stereotyping within a narrative that may be unattractive to certain groups of students. This seems to be especially true with regard to gender and ethnicity (Cooper, Hall, & Huff, 1990; Moreno & Flowerday, 2006). Certain themes in video games, such as competition and aggression, can lead to decreased interest and increased stress for females, especially in public situations (Cooper et al., 1990; Funk & Buchman, 1996b). Additionally, gender stereotyping and the lack of female representation in video games may be great contributing factors to the lower frequency of gameplay by females (Dietz, 1998). Student learning outcomes could be adversely affected since interest and anxiety have been found to be related to student domain knowledge and performance (Pekrun, 1992; Tobias, 1994). These prior studies on gender and video games suggest it is likely that males and females will experience flow differently for video games with differing narrative themes, and therefore have inconsistent learning gains as well.

This study considers the motivational benefits of the presence of a narrative in a math game and its relationship to gender and learning outcomes. In particular, the impact of the presence or absence of narrative was compared on student motivation and learning. Masculine and feminine narratives were developed to target any differences by student gender that may occur when they are given a narrative that is matched or mismatched to the student's gender. Previous studies considered narrative, motivation, and gender differences in commercial games, but the current study applied these concepts in a mathematics video game. We hypothesized that the presence of a narrative would yield increased learning and motivation and that there would also be gender effects based on the type of narrative the student received. Specifically we hypothesized:

- Hypothesis 1a: The presence of a narrative will increase students' motivation and game engagement;
- Hypothesis 1b: The presence of a narrative will increase student math learning;
- Hypothesis 2a: Students who are matched with the gender of the game avatar will have higher motivation and engagement compared to students who are mismatched with the gender of the avatar; and
- Hypothesis 2b: Students who are matched with the gender of the game avatar will have greater math learning compared to students who are mismatched with the gender of the avatar.

Game

In collaboration with CRESST, USC's Game Innovation Lab designed and developed a prototype game that was used as a research testbed (see Figure 1). The game requires students to apply concepts underlying rational number addition. Students are presented with the challenge of helping the game character Patch jump over hazards to safely reach a destination. To do so, students place trampolines at various locations along a one- or two-dimensional grid. Each trampoline is made bouncy by the student dragging one or more coils onto the trampoline. A coil has a value that represents a whole unit or fractions thereof. The distance Patch will bounce is the sum of all coils (values) added to the trampoline. For example, if the student drags a coil of 1 unit onto a trampoline, that trampoline will cause Patch to bounce exactly one unit.

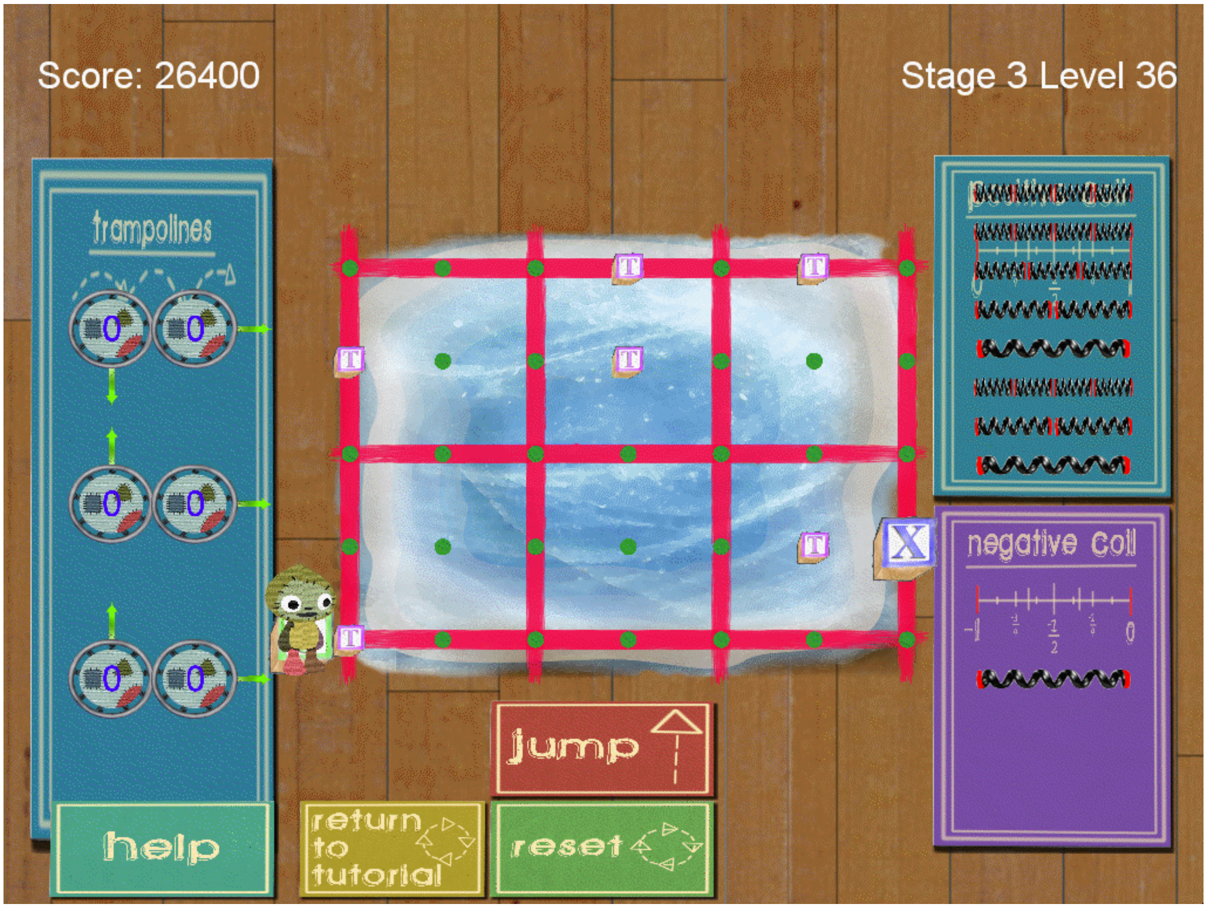


Figure 1. Screen shot of Save Patch.

In *Save Patch*, one whole unit is always the distance between two lines (Figure 1). This unit becomes the referent for coils of fractional bounce later on. Coils can be added to a trampoline to increase the distance Patch will bounce; however, only coils of the same fractional size can be added together. While any size coil can be placed on the trampoline initially, subsequent coils can only be added to the trampoline if they are the same size (i.e., have the same denominator).

The game exploits an important property of real numbers—numbers can be broken into smaller, identical parts to facilitate addition, a process that is similar in both integer and fractional addition. An important game design goal was to make the connections between integer addition and fractional addition explicit. Moreover, the gameplay requires that players focus on the size of a unit when they are adding coils to a trampoline.

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

In *Save Patch*, the procedure for converting fractions of different sizes (i.e., fractions with different denominators) is not accomplished through multiplication. Rather, students are shown how they could break the coil into more pieces (each smaller in size) or fewer pieces (each larger in size), respectively. For example, a coil that was one whole unit could be broken into two halves, three thirds, and so forth. If the student used the same procedure with a 1/2 coil, then the coil broke into two fourths, three sixths, and so on.

PILOT STUDY

Two narratives were tested in the pilot study with the main purpose of gathering information on how male and female students respond to and engage in games with different narrative themes and any impact this may have on learning outcomes.

Method

A pilot study was conducted to test measures and student responses to a narrative in the game. This study took place at a high school in Tulare, California.

Participants

Fifteen students enrolled in an Algebra 1 repeater class (taking Algebra 1 for at least the second time) participated in the pilot study. Nine participants were male (four 10th grade, five 11th grade) and six were female (all 10th grade). All students were Hispanic/Latino(a) and eight students reported speaking a language other than English all or most of the time at home. Students' math grades on their last report card ranged from C to F with nine receiving F's. The males reported one C, two D's, and six F's and females reported two C's, one D, and three F's.

Design

Two versions of the game were created, one with a feminine perspective and the other with a masculine perspective. Each game contained an avatar and narrative that matched the perspective of the game. Figure 2 shows the female and male avatars.

Figure 2. Female and male avatars.

Students were randomly assigned to conditions resulting in seven students in the male narrative condition and eight students in the female narrative condition. Table 1 gives the distribution of male and female students in the two conditions.

Table 1

Distribution of Males and Females by Narrative Condition

Gender	Female narrative	Male narrative
Male	5	4
Female	3	3

Measures

Math pretest and posttest. A 20-question (α = .59, *n* = 7) math pretest covering fractions was used to assess students' prior knowledge of rational numbers. The pretest included items on adding common and uncommon denominators, identifying pieces of a whole unit, and listing fractions by size. Students took a 29-item posttest (α = .88, *n* = 7), with the first 20 questions (α = .78, *n* = 8) covering topics similar to those on the pretest, and nine additional game-specific posttest items (α = .78, *n* = 8) specifically addressing understanding of math as presented in the game. Students having 20% of the pretest or posttest missing were not included in a calculation of the measure as these students were most likely noncompliant in completing the math measure and were not included in tests dealing with math outcomes.

Math attitudes. Students completed a 14-item survey adapted from Marsh, Hau, Artelt, Jurgen, and Peschar (2006) measuring their confidence in and attitudes toward math and school in general. From these items, scales were calculated to indicate students' self-belief and interest in math and academic tasks.

Game experience. Game experience was measured via self-reports of how much time participants spent playing games in general. Participants were instructed to indicate how often (a week) they played video games (1 = *none*, 2 = *1-2 hrs/week*, 3 = *3-6 hrs/week*, and 4 = *more than 6 hrs/week*).

Motivation and engagement outcomes. Students completed a survey after gameplay in order to measure perception of their experience with the game. A 4-point (1 = *disagree*, 2 = *somewhat disagree*, 3 = *somewhat agree*, 4 = *agree*) Likert scale was used for all items. Five scales were created from the items: *willingness to replay game, time distortion, negative* *perception of challenge*, *positive perception of challenge*, and *student evaluation of gaminess* (i.e., did the game incite the feelings of engagement and desire to play with friends that accompany video games). The *replay* scale consisted of nine items ($\alpha = .88$, $n = 14$) and were:

1. I became hooked on the game
2. I was concentrating a lot while playing the game
3. I enjoyed playing the game
4. I thought the game was fun
5. If the game had more levels, I would want to play them
6. I would play this game again
7. Beating the different levels made me feel good
8. I did not want to lose in the game
9. I cared about earning points in the game

There were five items ($\alpha = .88$, $n = 14$) in the *time distortion* scale and they were:

1. I was concentrating a lot while playing the game
2. It felt like I was playing the game for less time than I really did
3. I forgot about everything else around me while I was playing the game
4. Time seemed to go by quickly when I played the game
5. I lost track of time when playing the game

The *negative perceptions of challenge* scale consisted of five items ($\alpha = .48$, $n = 15$) and were:

1. I thought the game was frustrating
2. I thought the game was hard
3. I got annoyed playing the game
4. I was confused about how to play the game
5. I had to try really hard while playing the game

The scale *positive perceptions of challenge* consisted of five items ($\alpha = .85$, $n = 15$) and were:

1. I thought the game looked cool
2. I would have liked to play longer
3. I wish I had more time to play the game
4. I would play this game when I have free time

5. I liked that the game was hard sometimes

The *game features* scale consisted of three items ($\alpha = .77$, $n = 15$) and they were:

1. I lost track of time when playing the game
2. I really got into the game
3. I would tell my friends to play this game

Narrative survey. Students were asked to respond in writing to questions measuring their perception of the narrative version of the game such as their impression and enjoyment of the character and story in the game. The survey consisted of seven questions:

1. If you had a choice, which of the following avatars/characters would you choose? (the choices are presented in Figure 3)
2. What would the name of your character be?
3. Write a narrative (story) that would go along with the game you just played.
4. What grade level do you think your story would be appropriate for?
5. Did you like the character you were given in the game? Why or why not?
6. Was the story in the game interesting? Why or why not?
7. What grade level of student would think the story and game were interesting?

Figure 3. Four avatar choices given to students.

Responses to Questions 5 and 6 of the narrative survey were coded as either positive (e.g., “yes,” “of course,” or “definitely”) or negative (e.g., “no, it was boring”). Additionally, two students gave the response “not sure” for Question 6, which was coded as undefined.

Background information. Students were asked to provide background information such as year in school, gender, ethnicity, frequency using a language other than English at home, and prior math grades.

Results

Impressions of Characters and Narrative

A trend toward more students responding that they liked the character they were given in the game was observed for students matched with the gender of the avatar than those who were mismatched, $\chi^2(1, N = 15) = 2.02$, $p = .15$. Specifically, 13 of the 15 students responded that they liked the character they were given in the game, while only two said they did not like the character. Two students who disliked their character were mismatched with their gender. One was a male who received the female narrative and the other was a female who received the male narrative. This distribution is shown in Table 2:

Table 2

Distribution of Matched/Mismatched Males and Females by Whether They Liked the Character

Player relation with avatar	Liked character	
Match	Yes	No
Male		
Matched	4	0
Mismatched	4	1
Female		
Matched	3	0
Mismatched	2	1

Similarly, most students (12 out of 15) responded that they thought the story in the game was interesting. Of the students who did not respond positively to the narrative, one male who was mismatched with the gender of Patch said that the story was not interesting and two students (one a matched male and one a mismatched male) said “Not sure” or “Don’t know.” All female students, regardless of gender matching, responded positively to the game narrative. The distribution of responses regarding interest in the story is shown in Table 3 (see following page):

Table 3

Distribution of Matched/Mismatched Males and Females by Whether They Found the Story Interesting

Player relation with avatar	Found story interesting		
Match	Yes	No	Not sure
Male			
Matched	3	0	1
Mismatched	3	1	1
Female			
Matched	3	0	0
Mismatched	3	0	0

Math Outcomes

Although the difference was not significant, overall posttest scores were higher for students who were matched (n = 5, M = 10.3, SD = 6.1) with the gender of Patch than for those who were mismatched (n = 4, M = 7.7, SD = 2.4). Mismatched students (n = 7, M = 8.0, SD = 3.7) scored only slightly lower than matched students (n = 5, M = 8.2, SD = 2.5) on the pretest. The non-significance observed here may be an artifact of small sample size especially given the moderate effect size, d = 0.61.

The means for scales of motivation and engagement from the narrative pilot were compared to those collected from another study with no narrative version. The study took place at the same school in an Algebra 1 class taught by the same teacher used for the narrative pilot. Twenty-four students participated in the study with no narrative implementation. The purpose of this comparison was to see if students would respond more positively to the game when they played a narrative version. All positive measures of student engagement, motivation, and enjoyment were higher for students in the narrative pilot. These measures included: student willingness to replay the game, students' feelings of time distortion while playing, a positive perception of game challenge, and perception of game features. Differences were significant for students' positive perception of challenge, $t(36)$ = 2.55, p = .015, and perception of game features, $t(36)$ = 2.02, p = .05. Table 4 presents the descriptive statistics for the common scales between the second Algebra 1 class and the narrative pilot. Effect sizes range from small to large; Cohen's d values are listed in Table 4 (see following page):

Table 4

Descriptives for Scales From Tulare Study and Overall Narrative Pilot

	No narrative version			Narrative pilot			Effect size
Scales	*n*	*M*	*SD*	*n*	*M*	*SD*	Cohen's *d*
Replay game	22	2.62	0.81	14	2.99	0.55	0.53
Time distortion	23	2.70	0.85	14	2.91	0.85	0.25
Positive challenge	26	2.35	0.78	15	3.00	0.76	0.86
Negative challenge	24	2.14	0.54	15	2.21	0.51	0.14
Game features	23	2.26	0.81	15	2.82	0.88	0.69

Discussion

Students who participated in the narrative pilot indicated high levels of engagement, interest, and enjoyment of the game. Mean responses to scales measuring desire to replay, perception of time distortion, positive perception of challenge, and perception of game features ranged from 2.8 to 3.0; items used a 4-point scale (1 = *disagree*, 2 = *somewhat disagree*, 3 = *somewhat agree*, 4 = *agree*) so a mean of 3.0 indicates that students responded at the positive end. In fact, the median score for these scales ranged from 2.94 to 3.0. A comparison of scale means for the pilot study with a similar class of students playing a version of the game without a narrative suggests that the presence of the narrative increased motivation and enjoyment of the game. However, other differences between the two studies may account for some of these differences as well; a non-narrative version needed to be included in the main study as a control. This trend is interesting and forms the basis for one hypothesis for the main study: that motivation, engagement, and interest in the game will be higher for the narrative versions of the game.

We are particularly interested in any learning gains that may be caused by the implementation of a narrative in a math game. Higher posttest scores for students who were matched with the gender of Patch than for mismatched students suggested that learning outcomes may be impacted by gender matching. This difference was not significant but the moderate Cohen's *d* suggested an effect of gender matching. For this reason, the main study also included a male and female narrative so that student performance based on gender could be considered between the two.

MAIN STUDY

Method

Design

A 3-group pretest/posttest design contrasted two story frames and a control condition. The two types of narrative themes were (a) masculine (aggressive, focused on a final showdown fight), and (b) feminine (focused on the relationship of twin sisters). The control condition had no narrative and the gender of the avatar was unspecified.

Participants

One hundred seventy-seven students in Grades 6-12 participated in this study. Students were enrolled in one of the following courses: 6th grade math, pre-algebra, algebra, or Algebra 1 repeater. Participants were randomly assigned to one of three conditions resulting in 62 participants in the masculine game narrative condition, 68 participants in the feminine narrative condition, and 47 participants to the control condition (i.e., no narrative). Seventy-six percent of students were in Grades 8 to 10. The distribution of students by grade in school is listed in Table 5.

Table 5

Frequency and Percentage of Participants by Grade

Grade	Number of students	Percentage
6	11	6.2
7	11	6.2
8	22	12.4
9	76	42.9
10	37	20.9
11	10	5.5
12	2	1.2
Total	169	95.5

Note. Eight students did not report their grade in school.

Eighty-five males and 80 females participated in the study, and 12 students did not report their gender. Table 6 shows the distribution of each of the three conditions by gender;

there were no significant differences for measures of math knowledge between the six groups.

Table 6

Distribution of Males and Females for Study Conditions

	Narrative theme			
Gender	None	Masculine	Feminine	Total
Male	23	30	32	85
Female	19	27	34	80
Total	42	57	66	165

Note. Twelve students did not report their gender.

Other demographic data included prior math grades, ethnicity, and frequency speaking a language other than English at home. On their last report card, 17.5% of students reported A's, 16.9% B's, 14.7% C's, 18.6% D's, and 20.9% F's for math. Eleven percent of students did not report their last math grade. Over 50% of students were Latino/a, 15% were White, and all other groups (African American, Asian or Pacific Islander, and multiethnic or biracial) were below 7% each. Approximately 45% of students reported speaking a language other than English outside of the home half or more of the time.

Game and Narrative Themes

The narrative theme was presented to players via a series of images that popped up at specific points in the game to further the story as the player completed certain portions of the game. Both versions had similar pacing and the masculine-themed game had 11 story images and the feminine-themed game had 10. The images and text contained in them were made feminine and masculine based on prior video game research on sex-stereotyping in game software. Cooper et al. (1990) found that when given a male-oriented computer game, females experienced more stress during gameplay than when they played the female-oriented game, especially in a public setting, and the same was true for male students when playing the female-oriented game. Male players tend to prefer games that involve violence, action, and aggression while females enjoy whimsical themes that emphasize cooperation without aggressive action (Cooper et al., 1990; Funk & Buchman, 1996a; Malone, 1981; Morlock, Yando, & Nigolena, 1985). For this reason, the main difference in the goal within the feminine and masculine narratives was the presence of aggression or emphasis on a final showdown in the form of a fight in the masculine narrative. Both stories introduced the antagonist, the Evil Skull Puppet, and the goal in the masculine and feminine versions was to get across his traps for an express purpose. In the male version of the game, the story emphasized a need to fight and conquer the Evil Skull Puppet because he had taken over the kingdom of his stepbrother, the protagonist Patch. Figure 4 shows an example of a screen from the male narrative of the game.

Figure 4. Example of narrative screen from masculine-themed version of the game.

In the female version of the game, the player must also reach the Evil Skull Puppet's castle but the player must do so to save Patch's twin sister who has been taken captive by the Evil Skull Puppet. Unlike the male version of the game, there is no reason to fight the Evil Skull Puppet and the story focuses on the relationship between the sisters, not on the animosity Patch has toward the Evil Skull Puppet (see Figure 5 for a sample screen from the female narrative). All images for both versions of the game and when they appear in the game are included in Appendix A.

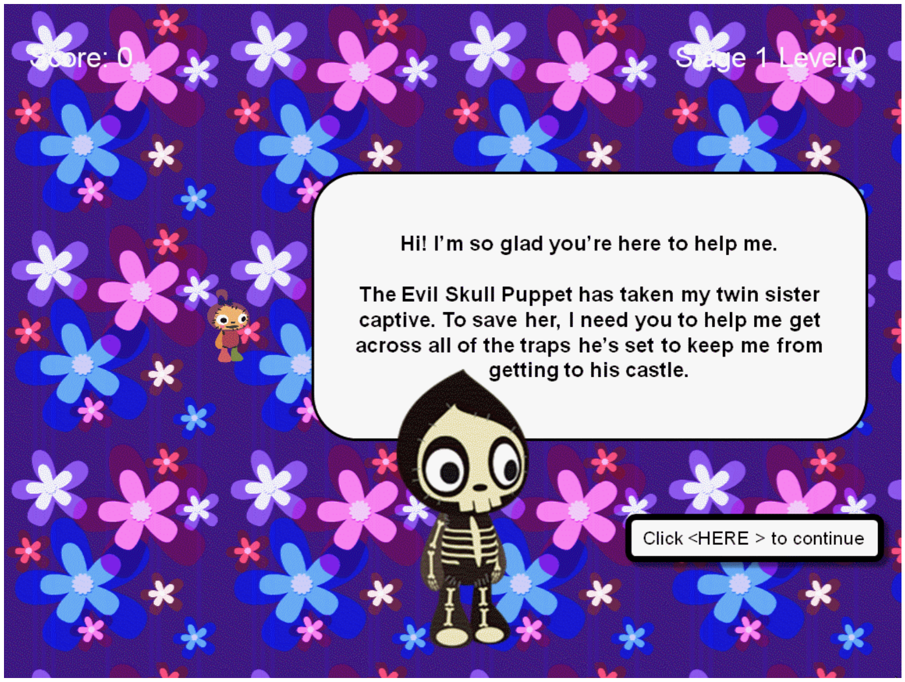


Figure 5. Example of narrative screen from feminine-themed version of the game.

Measures

Math Knowledge

A 23-question (α = .89, *n* = 115) math pretest was used to assess students' prior knowledge. The pretest content included adding common and uncommon denominators, identifying pieces of a whole unit, and listing fractions by size. After gameplay, students took a 34-item (α = .92, *n* = 115) posttest which included 24 questions (α = .89, *n* = 106) similar to those on the pretest and included 10 additional items (α = .90, *n* = 143) specifically addressing understanding of math as used in the game. Two exploratory measures of change were used: Normalized change (Marx & Cummings, 2007) and symmetricized percent change (SPC) (Berry & Ayers, 2006).

Normalized change was defined as $\frac{posttest\%correct - pretest\%correct}{100 - pretest\%correct}$, with adjustments made to the computation for the cases when pretest % correct is 0% or 100% as suggested by Marx and Cummings (2007). The range of normalized change score is –100% to +100%. SPC was defined as $100 \times \frac{posttestscore - pretestscore}{posttestscore + pretestscore}$ and is 0 when the posttest score is equal to the pretest score, with a possible range of –100% to +100%.

Motivation and Engagement Survey

Scales were created from items targeting students' motivation and engagement while playing the game. All scales were validated using confirmatory factor analysis. A 4-point (1 = *disagree*, 2 = *somewhat disagree*, 3 = *somewhat agree*, 4 = *agree*) Likert scale was used for all items. For example, a 10-item scale of motivation to replay the game was created from items asking students to rank their desire to reach the next level, to help the character, and to find out what was going to happen, and their interest in the game (α = .94, *n* = 160). The following questions were included in the *replay* scale:

1. I would have liked to play longer
2. I wish I had more time to play the game
3. I would play this game when I have free time
4. I would tell my friends to play this game
5. If the game had more levels, I would want to play them
6. I would play this game again
7. I thought the game looked cool
8. I liked that the game was hard sometimes
9. I enjoyed playing the game
10. I thought the game was fun

Another scale of interest was created from four items measuring negative feelings about the challenge in the game (α = .70, *n* = 170); questions included in this scale were:

1. I thought the game was frustrating
2. I got annoyed playing the game
3. Playing the game was boring
4. I was confused about how to play the game

A seven-item scale was created to measure students' perception of flow in the game (α = .89, *n* = 163). The following questions were used to form the *perception of flow* scale:

1. I really got into the game
2. I was concentrating a lot while playing the game
3. I lost track of time when playing the game
4. I forgot about everything else around me while I was playing the game
5. Time seemed to go by very quickly when I played the game
6. It felt like I was playing the game for less time that I really did

7. I became hooked on the game

Narrative Survey

The following four questions (α = .84, n = 176) were used to create a *narrative interest* scale:

1. I was motivated to find out what was going to happen to Patch
2. I thought Patch was an interesting character
3. The game was interesting to me
4. I was excited to reach the next level as I was playing the game

Seven items (α = .82, n = 169) were used to create a *game self-efficacy* scale:

1. I understood how to move forward in the game.
2. I really understood how to help Patch
3. I understood how to play the game
4. I think the game made my math skills better
5. I learned something new about fractions from the game
6. If someone asked me, I could explain what the green dots mean
7. I knew what the goals of the game were

Finally, an open-ended question asked the student to explain why they did or did not care about reaching the end. This item was coded to account for positive reasons for wanting to reach the end of the game such as an interest in the narrative or the character, as well as negative reasons (e.g., bored and just wanted the game to end). Negative reasons for wanting to reach the end of the game were placed in the same group as students who were not interested in reaching the end of the game. The open-ended question was scored to account for the presence or absence of certain themes in their explanations. The eight themes of interest were: positive explanations, negative explanations, positive opinion of the game character, negative opinion of the game character, positive impression of the story, negative impression of the story, positive feelings about the math, and negative feelings about the math. A rubric was created to guide raters as they evaluated each response for the presence or absence of each variable. A training set of five representative responses had 100% agreement between two raters for each variable. Thirty-six responses (20%) were selected for scoring by both raters and reliability ranged from .83 to 1.0, kappa value. Appendix B displays the rubric used for scoring each variable and the associated kappa.

Background Survey

Students were asked to provide background information such as gender, prior math grades, and amount of time spent playing video games. These surveys are listed in Appendix C and Appendix D respectively.

Results

Descriptives

Comparisons were made between gender and conditions prior to analysis to ensure that there were no differences that would impact learning outcomes. A *t* test showed that there was no significant difference between the performance of males (M = 14.2, SD = 6.14) and females (M = 14.9, SD = 6.04) on the math pretest. There was also no significant difference between males' and females' performance on the math posttest or on the game-specific conceptual math questions. No significant differences existed among the three conditions for student performance on the math pretest. A score was not calculated for a student missing 20% or more of the questions that formed each math scale. Of the 177 participants, 165 completed at least 80% of the math pretest, 162 completed at least 80% of the math posttest, and 154 students completed at least 80% of the game-specific conceptual items. Descriptive statistics for these and other measures used in this study are listed in Table 7. Additionally, descriptive analyses for the missing data are included in Appendix E.

Table 7

Descriptives for Study Measures

Descriptives	*n*	*M*	*SD*	Min.	Max.	Cronbach's alpha
Math knowledge						
Math pretest	167	11.67	5.13	2.50	22.17	.905[a]
Math posttest	160	11.82	5.12	3.00	23.00	.894[a]
Game posttest	155	5.21	2.70	.00	10.00	.909[b]
Symmetric change	156	-.04	8.88	-28.57	20.00	---
Normalized change	156	.03	.22	-.44	1.00	---
Game engagement						
Replay game	160	2.50	.89	1.00	4.00	.944[b]
Negative challenge perception	170	2.34	.79	1.00	4.00	.701[c]
Game flow	163	2.57	.83	1.00	4.00	.893[d]
Narrative perceptions						
Narrative interest	176	2.65	.91	1.00	4.00	.842[c]
Game self-efficacy	169	2.87	.66	1.00	4.00	.823[d]

[a]Number of items = 23. [b]Number of items = 10. [c]Number of items = 4. [d]Number of items = 7.

Correlation

Pearson's correlations were calculated for math knowledge, game engagement, narrative perception, and background variables. Table 8 lists these correlations.

Table 8

Relation of Matched Student and Avatar Gender with Learning Outcomes and Game Perception

Measure	1	2	3	4	5	6	7	8	9	10	11
Background											
1. Freq. gameplay	---										
2. Last year's math grade	.20*	---									
Math knowledge											
3. Math pretest	.13	.44***	---								
4. Math posttest	.05	.45***	.95***	---							
5. Game posttest	.07	.34***	.81***	.81***	---						
6. Symmetrical change	-.04	-.00	-.11	.19*	.12	---					
7. Normalized change	-.06	.24***	.30***	.55***	.41***	.69***	---				
Game engagement											
8. Replay game	-.06	.03	.08	.05	.22***	-.01	.09	---			
9. Negative challenge perception	-.08	-.10	-.13	-.12	-.19*	.12	-.02	-.56***	---		
10. Game flow	-.07	.06	.05	.00	.17*	-.07	-.04	.83***	-.41***	---	
Narrative perception											
11. Narrative interest	-.04	.15	.08	.05	.23***	-.08	.05	.81***	-.50***	.69***	---
12. Game self-efficacy	-.01	.12	.23***	.15	.36***	-.01	.09	.73***	-.42***	.65***	.69***

*p < .05, two-tailed. **p < .01, two-tailed. ***p < .001, two-tailed.

Background information. Somewhat surprisingly, frequency of video game play was found to have a significant positive relationship with the students' reported math grades from the previous year, $r(146) = .20$. This suggests that students who played video games more frequently also had higher math grades during the previous year. Frequency of gameplay was not related to any other variables, but previous math grades had significant positive correlations with all of the math knowledge variables except for symmetric change.

Learning outcomes. Aside from the positive correlation of several of the math knowledge scales with prior math grades, the majority of math scales were also positively correlated with each other at p values < .001. The outcome measure symmetric change is the exception and is only related to posttest at a .05 level, $r(156) = .19$, and normalized change at the .001 level, $r(156) = .69$.

There were also relationships between the math knowledge variables and game engagement and narrative perception scales. Pretest was positively related to students' game self-efficacy, $r(159) = .23$, implying that students with higher pretest scores were more likely to report understanding of the game. The game conceptual posttest items were also positively correlated with game self-efficacy, $r(149) = .36$, narrative interest, $r(154) = .23$, willingness to replay, $r(143) = .22$, and game flow perception, $r(143) = .17$. Conceptual items were negatively related to students' negative perception of game challenge, $r(150) = -.19$, which indicates the higher a student scored on the conceptual game items, the more likely they were to have enjoyed the challenge presented in the game. Other posttest items, symmetric change, and normalized change did not have any significant relationships with the engagement and perception scales.

Game engagement and narrative perception. All variables measuring favorable engagement in the game and favorable perception of the game narrative were positively related to each other at the .001 level. Negative perception of game challenge was negatively correlated with all of these scales at the .001 level. The positive relationship of game self-efficacy and the other engagement and perception variables suggests that the more a student enjoyed the game and narrative, the more likely he or she felt that they understood how to play the game.

Group Differences

A series of planned contrasts were performed to test our four hypotheses. As secondary analyses, linear regression and covariate analysis were conducted.

Hypothesis 1a: The presence of narrative will increase students' motivation and game engagement.

Contrast tests. There were no significant differences between the masculine and feminine narratives on any of the game engagement or narrative perception scales. This allowed for the pooling of the two narrative conditions so that contrast tests between narrative and no narrative versions could be analyzed. Game flow was the only planned contrast test that confirmed the hypothesis of increased engagement when given a narrative $t(161) = 2.0$, $p = .05$. Contrast tests were run for replay, negative perception of game challenge, narrative interest, and game self-efficacy but these tests did not confirm the hypothesis. Game self-efficacy was not significant with $t(167) = 1.4$, $p = .16$. Effect sizes (Cohen's *d*) were 0.32 and 0.24 for game flow and game self-efficacy, suggesting a practical difference between narrative and no narrative. Higher perceived understanding of the game and the concepts it was trying to teach by students receiving a narrative version of the game is also consistent with results from the study by Koenig (2008).

Hypothesis 1b: The presence of narrative will increase students' math learning.

Contrast tests. Students in the masculine narrative condition ($n = 53$, $M = .07$) obtained significantly higher normalized change scores than students who received the feminine narrative ($n = 60$, $M = -0.02$), $t(111) = 2.2$, $p = .03$. Since there was a significant difference between the two narrative conditions, they could not be pooled and a contrast test was not conducted to compare narrative and no narrative for normalized change. Math posttest, game conceptual posttest, and symmetric change were not significantly different for the masculine and feminine narratives so they were pooled and contrast tests conducted. However, there were no significant differences between narrative and no narrative, so our original hypothesis was not confirmed.

Conceptual posttest differences. There were marginally significant condition differences observed on game-specific conceptual math questions when an ANCOVA was performed with pretest as a covariate, $F(2, 146) = 3.00$, $p = .053$, eta = .040. Post hoc LSD analysis shows that students in the no narrative condition (adj. $M = 7.9$, $SE = .58$) scored lower on the conceptual math questions than those who received the male narrative (adj. $M = 9.4$, $SE = .51$), with a marginal significance of $p = .064$. Those who received the female narrative (adj. $M = 7.8$, $SE = .47$) also scored lower than those receiving the male narrative with $p = .023$. There appears to be no trend toward a difference between students with no narrative and female narrative conditions, $p = .845$.

Male narrative differences. To consider the possible increase in learning outcomes for students in the masculine narrative condition, a *t* test was performed to compare male narrative to the other two conditions. A *t* test comparing feminine narrative and the no narrative condition confirmed that the two conditions could be combined as there were no significant differences for any of the learning outcome measures. For the normalized change measure, the students in the masculine narrative condition (n = 53, M = .07) exhibited more change from the pre- to posttest than students not in the masculine narrative condition (n = 103, M = .002); this difference was marginally significant $t(154) = -1.85$, $p = .066$.

Students in the masculine narrative condition also appeared to score higher than those in the other conditions on the game-specific conceptual math posttest items. Linear regression analysis revealed that the pretest and whether the student received the masculine narrative predicted performance on the conceptual posttest items. Participants in the masculine narrative condition scored 0.848 points higher on the conceptual items, $\beta = 0.848$, $t(138) = 2.42$, $p < .017$, and for every point higher a student scored on the pretest, their conceptual score increased by 0.424 points, $\beta = 0.424$, $t(132) = 16.42$, $p < .001$. The same model was run for normalized change, symmetric change, and overall posttest score but only pretest was a significant predictor.

Hypothesis 2a: Students who are matched with the gender of the game avatar will have higher motivation and engagement compared to students who are mismatched with the gender of the avatar.

Contrast tests. Our hypothesis could not be confirmed by planned contrast tests comparing game engagement and narrative perception between students matched or mismatched with the gender of the avatar they received in the game.

Matching of male students. Cross-tabulation comparing males who were matched with the gender of the game avatar and those who were not matched males (i.e., matched females, mismatched males, and no gender specification) indicated that matched males reported more frequently positive reasons for being motivated to reach the end of the game than those who were not matched males, χ^2 (1, n = 122) = 4.88, p = .04. The distribution of responses is shown in Table 9; 61 of 101 (60%) non-matched male students gave a positive explanation for wanting to reach the end of the game, while 18 of 21 (85%) of matched males cited positive reasons.

Table 9

Distribution of Positive Responses for Desire to End Game

Gender Matching	Positive responses: Yes	No	Total
Matched male	18	3	21
Not matched male	61	40	101
Total	79	43	122

With the exception of three students, all students who did not list a positive reason for wanting to reach the end of the game gave negative reasons for wanting to reach the end. An example of this type of response is, "I was so bored, I couldn't wait for the game to end." Table 10 shows that three non-matched male students did not address negative or positive reasons for wanting to reach the end of the game. Three of 21 (14%) matched male students gave negative reasons for wanting to reach the end of the game, while 37 out of 101 (37%) of the non-matched male students had negative responses. This difference was marginally significant, χ^2 (1, n = 122) = 3.94, p = .07.

Table 10

Distribution of Negative Responses for Desire to End Game

Gender Matching	Negative responses: Yes	No	Total
Matched male	3	18	21
Not matched male	37	64	101
Total	40	82	122

Hypothesis 2b: Students who are matched with the gender of the game avatar will have greater math learning compared to students who are mismatched with the gender of the avatar.

Contrast tests. Differences in math learning outcomes were also not significant for contrast tests comparing students who were matched and students who were mismatched with the gender of the game avatar.

Matching of male students. In a linear regression model, pretest and the matching of male students predict conceptual item scores. For every point higher a student scored on the pretest they were expected to score 0.42 points higher on the conceptual posttest items, $\beta = 0.420$, $t(143) = 16.58$, $p < .001$. A matched male student was expected to score approximately one point higher on the conceptual items than those who were not matched males, $\beta = 0.955$, $t(143) = 2.51$, $p < .013$. Matching of female students was also included in the model, but did not predict significant change in conceptual posttest scores.

Discussion and Significance

Narrative (Hypothesis 1)

Motivation and engagement (Hypothesis 1a). Results provide evidence to support the hypothesis that students receiving a narrative version of a learning game would be more engaged in the game than those receiving no narrative. Contrast testing confirmed that students with the narrative versions of the game experienced more game flow, and presumably were more engaged in the game as experiencing flow results in loss of conception of time and focus on a task (Csikszentmihalyi, 1991; Jennings, 2002). Other measures of game engagement did not show differences among narrative and no narrative versions of the game. However, this is not entirely surprising since the three games were exactly identical except for the narrative content and the experience of flow has been used as a measure of increased engagement in a task with a narrative for several studies (Green et al., 2004; Park et al., 2010; Schneider et al., 2004).

Learning outcomes (Hypothesis 1b). Although learning gains did occur between narrative and non-narrative versions, this was confounded by differences in learning that also occurred between students receiving the male and female narratives. Consistency with prior research indicating that increased experience of flow also results in increased learning outcomes would suggest that both narrative versions of the game would experience higher learning gains than the no narrative version. There was no difference between the two narrative versions in experience of flow; however, learning gains were only experienced by students in the masculine narrative condition. This was true for two measures of learning outcome: normalized change score and performance on game-specific conceptual items. The reason for this difference is unclear; however, the implementation of a feminine narrative in a video game is uncommon, which may distract from the benefit of the inclusion of a narrative (Dietz, 1998; Kafai, 1996; Subrahmanyam & Greenfield, 1998). In any case, more research is needed to examine the role of feminine versus masculine narratives in learning games.

Gender (Hypothesis 2)

Motivation and engagement (Hypothesis 2a). Analysis of open-ended items measuring students' reasons for wanting to reach the end of the game indicated that only males matched with the gender of the game avatar and narrative reported increased positive motivations. There seemed to be no impact on motivation when females were matched to the avatar and narrative or for those students who did not specify their gender. This result seems contrary to findings indicating that females do not play video games as frequently as males due to the lack of female representation and to gender stereotyping in games (Cooper et al., 1990; Funk & Buchman, 1996b; Moreno & Flowerday, 2006). Males tend to be the most frequent video game players and the success of the video game industry has depended largely on males purchasing video games with themes that generally appeal more to males than females (Dietz, 1998; Hartmann & Klimmt, 2006; Kafai, 1996, 1998; Subrahmanyam & Greenfield, 1998). Our results seem to confirm that it is the perception of males that are more affected by the use of narratives in video games. Reasons for this may be that female students are already less interested in the game even if they receive the feminine narrative simply based on preconceptions of video game play or the feminine narrative may have been insufficient in targeting the themes that female students find engaging.

Learning outcomes (Hypothesis 2b). Males matched with the gender of the game avatar and narrative themes also experience learning gains, measured by performance on game conceptual posttest items, while the same was not true of matched females. This is consistent with results from Hypothesis 2a and narrative engagement research; matched females did not report increased engagement or experience learning gains, while matched males had both.

Conclusions

This study indicates that the use of narrative in educational video games has the potential to increase student engagement and learning outcomes, especially when males are provided with a male avatar and masculine narrative. Overall, the presence of a narrative resulted in increased flow regardless of gender; however, learning gains only occurred for students given the masculine narrative. This may be confounded by the fact that matched male students experienced the most increase on learning outcomes. Further research in this area should focus on the potential of narrative or other components of game design to be used to increase learning for females in educational games.

REFERENCES

Berry, D. A., & Ayers, G. D. (2006). Symmetrized percent change for treatment comparisons. *American Statistical Association, 60*, 27–31.

Bielenberg, D. R., & Carpenter-Smith, T. (1997). Efficacy of story in multimedia training. *Journal of Network and Computer Applications, 20,* 151-159.

Busselle, R., & Bilandzic, H. (2008). Fictionality and perceived realism in experiencing stories: A model of narrative comprehension and engagement. *Communication Theory, 18,* 255-280.

Cooper, J., Hall, J., & Huff, C. (1990). Situational stress as a consequence of sex-stereotyped software. *Personality and Social Psychology Bulletin, 16,* 419-429.

Csikszentmihalyi, M. (1988). The flow experience and its significance for human psychology. In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.), *Optimal experience: Psychological studies of flow in consciousness* (pp. 15 – 35). Cambridge, UK: Cambridge University Press.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience*. New York, NY: Harper Perennial.

Dietz, T. L. (1998). An examination of violence and gender role portrayals in video games: Implications for gender socialization and aggressive behavior. *Sex Roles, 38,* 425-442.

Funk, J. B., & Buchman, D. D. (1996a). Children's perceptions of gender differences in social approval for playing electronic games. *Sex Roles, 35*, 219-231.

Funk, J. B., & Buchman, D. D. (1996b). Playing violent video and computer games and adolescent self concept. *Journal of Communication, 46*, 19-32.

Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment: The role of transportation into narrative worlds. *Communication Theory, 14,* 311-327.

Hartmann, T., & Klimmt, C. (2006). Gender and computer games: Exploring females' dislikes. *Journal of Computer-Mediated Communication, 11,* 910-931.

Jennings, A. S. (2002). Creating an interactive science murder mystery game: The optimal experience of flow. *IEEE Transactions on Personal Communication, 45*, 297-301.

Kafai, Y. B. (1996). Gender differences in children's construction of video games. In P. M. Greenfield & R. R. Cocking (Eds.), *Interacting with video* (pp. 39-66). Norwood, NJ: Ablex Publishing.

Kafai, Y. B. (1998). Video game designs by girls and boys: Variability and consistency of gender differences. In J. Cassell & H. Jenkins (Eds.), *From Barbie to Mortal Combat* (pp. 46-71). Boston, MA: The MIT Press.

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and Higher Education, 8*, 13-24.

Koenig, A. (2007, July). *Exploring narrative as a cognitive aid in educational video games*. Paper presented at the 13th International Conference on Artificial Intelligence in Education (AIED), Marina Del Rey, CA.

Koenig, A. D. (2008). *Exploring effective educational video game design: The interplay between narrative and game-schema construction* (Unpublished doctoral dissertation). Arizona State University, Department of Educational Technology.

Konradt, U., & Sulz, K. (2001). The experience of flow in interacting with a hypermedia learning environment. *Journal of Educational Multimedia and Hypermedia, 10,* 69-84.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 4,* 333-369.

Marsh, H. W., Hau, K. T., Artelt, C., Jurgen, B., & Peschar, J. L. (2006). OECD's brief self-report of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing, 6*(4), 311-360.

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

Moreno, R., & Flowerday, T. (2006). Students' choice of animated pedagogical agents in science learning: A test of the similarity attraction hypothesis on gender and ethnicity. *Contemporary Educational Psychology, 31,* 186-207.

Morlock, H., Yando, T., & Nigolena, K. (1985). Motivation of video game players. *Psychological Reports, 57,* 247-250.

Park, N., Lee, K. M., Jin, S. A., & Kang, S. (2010). Effects of pre-game stories on feelings of presence and evaluation of computer games. *International Journal of Human-Computer Studies, 68,* 822-833.

Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology: An International Review, 41,* 359-376.

Schneider, E. F., Lang, A., Shin, M., & Bradley, S. D. (2004). Death with a story: How story impacts emotional, motivational, and physiological responses to first-person shooter video games. *Human Communication Research, 30,* 361-375.

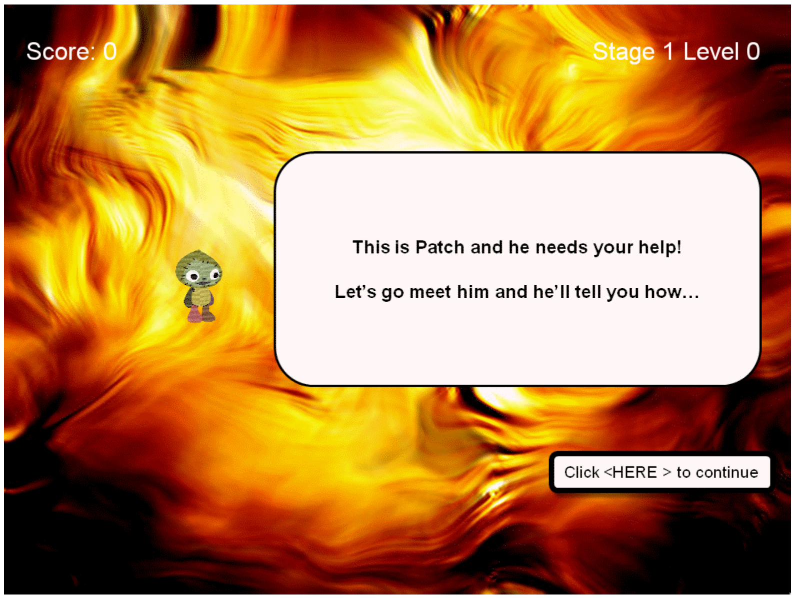
Skadberg, Y. X., & Kimmel, J. R. (2004). Visitors' flow experience while browsing a web site: Its measurement, contributing factors, and consequences. *Computers in Human Behavior, 20,* 403–422.

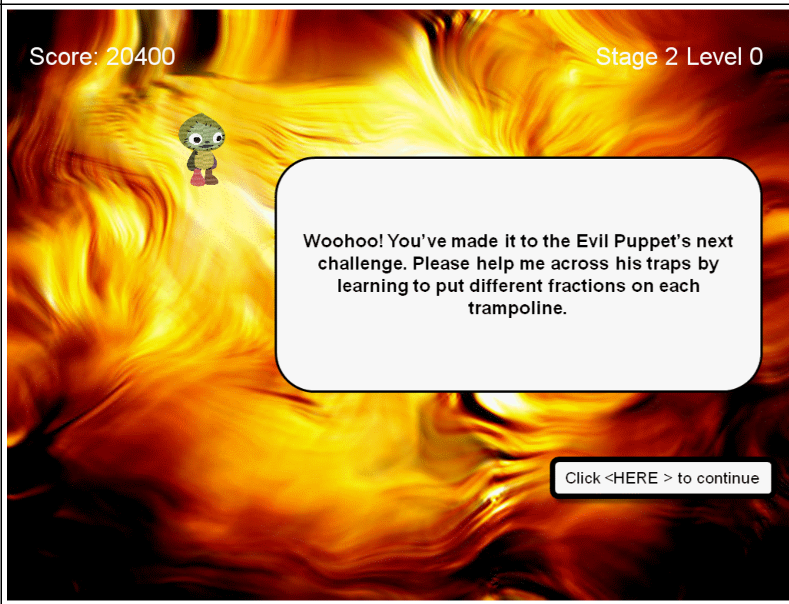
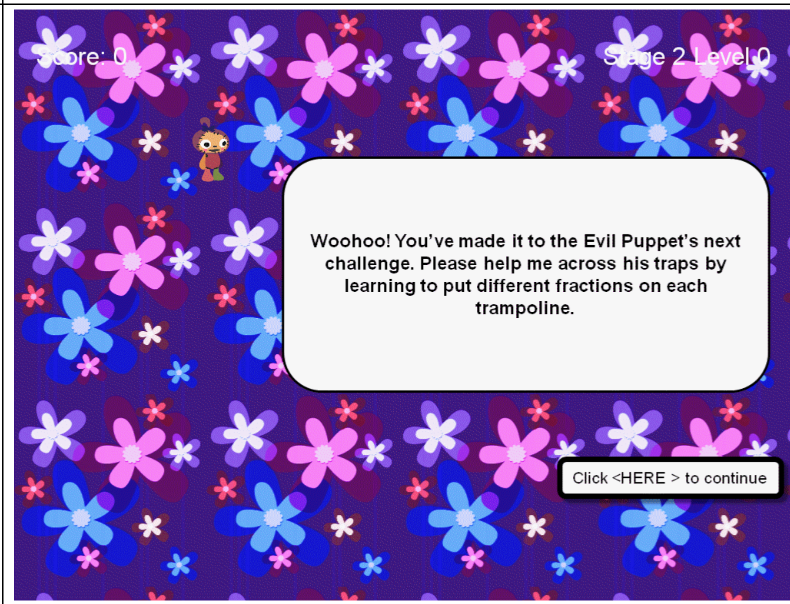
Subrahmanyam, K., & Greenfield, P. M. (1998). Computer games for girls: What makes them play? In J. Cassell & H. Jenkins (Eds.), *From Barbie to Mortal Combat* (pp. 46-71). Boston, MA: The MIT Press.

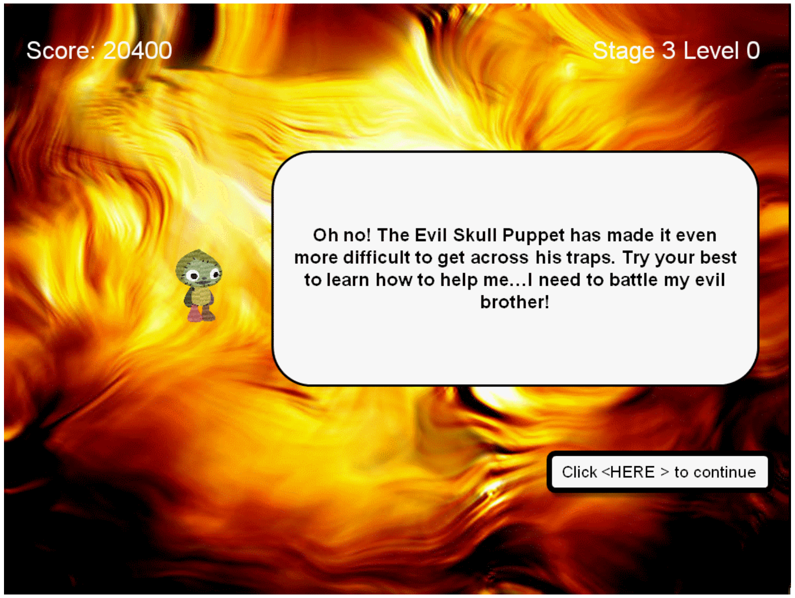
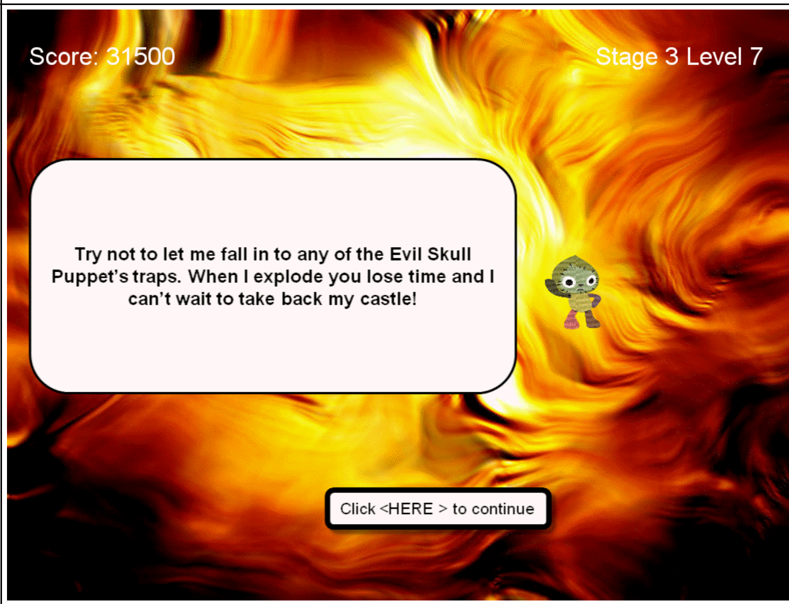
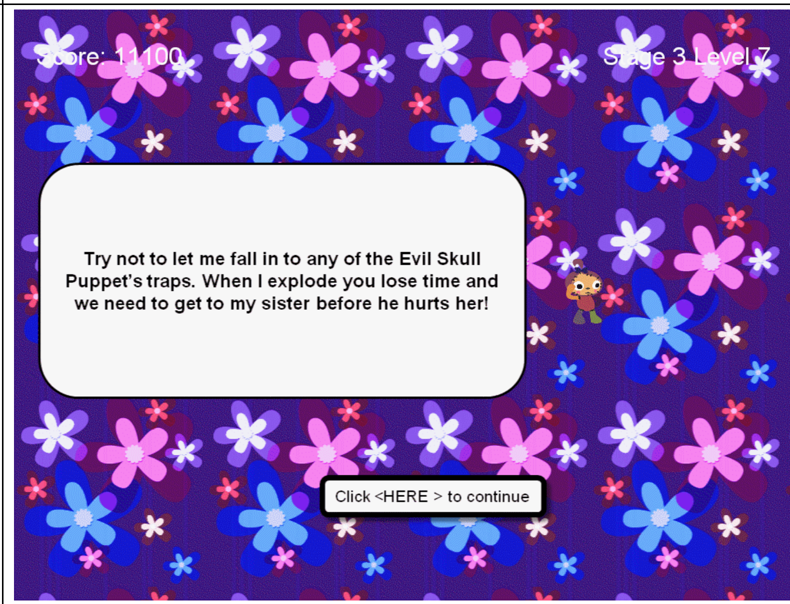
Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64,* 37-54.

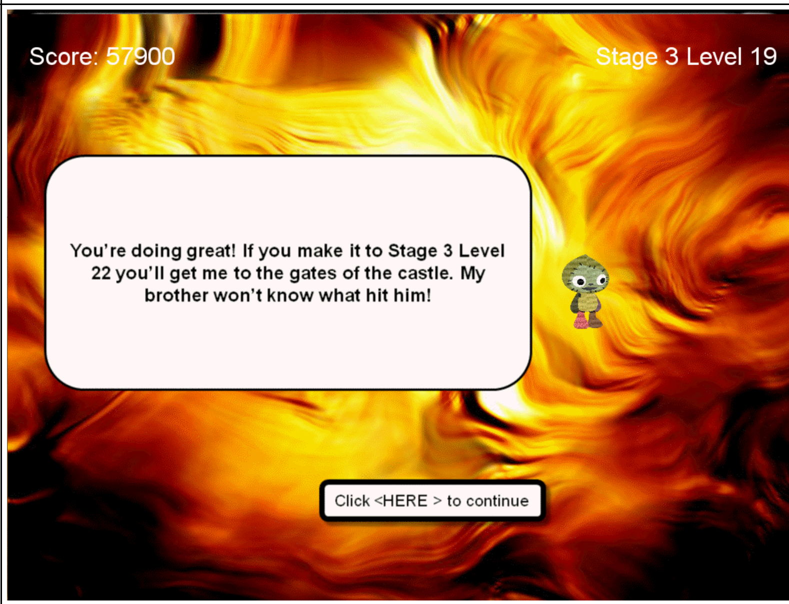
Webster, J., Trevino, L. K., & Ryan, L. (1993). The dimensionality and correlates of flow in human-computer interaction. *Computers in Human Behavior, 9,* 411-426.

APPENDIX A
NARRATIVE SCREENS

Level	Masculine narrative	Feminine narrative
1-0		

<table>
<tr><th>Level</th><th>Masculine narrative</th><th>Feminine narrative</th></tr>
<tr><td>Start 1-0</td><td>
</td><td>
</td></tr>
<tr><td>Start 2-0</td><td>
</td><td>
</td></tr>
</table>

<table>
<tr><th>Level</th><th>Masculine narrative</th><th>Feminine narrative</th></tr>
<tr><td>Start 3-0</td><td>
</td><td>
</td></tr>
<tr><td>End 3-7</td><td>
</td><td>
</td></tr>
</table>

Level	Masculine narrative	Feminine narrative
End 3-13		
End 3-19		

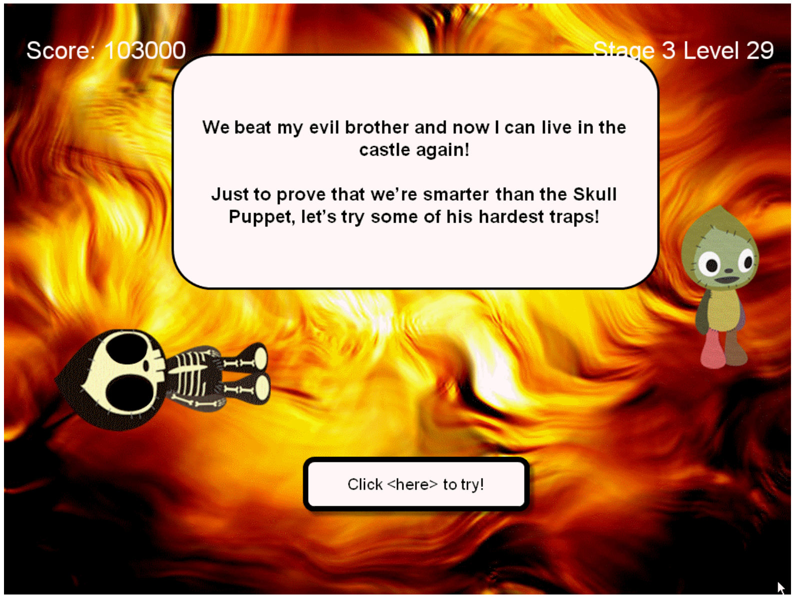
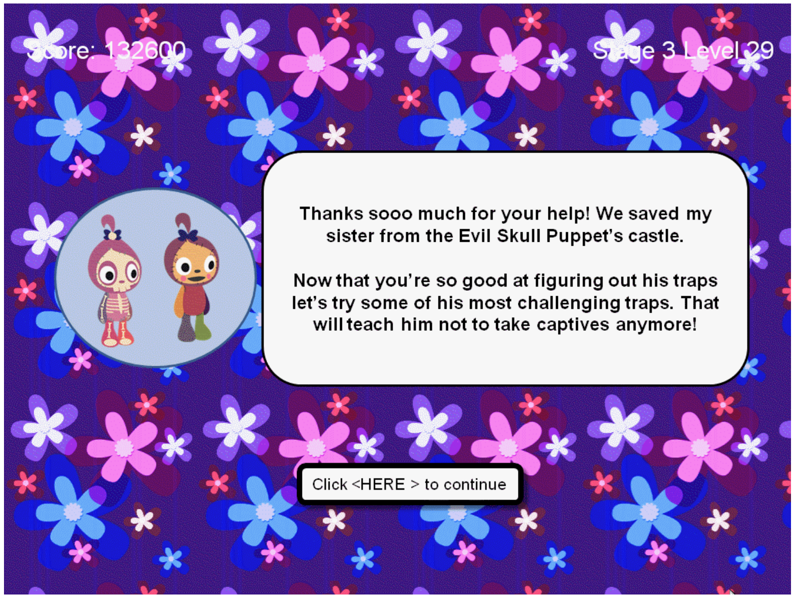
<table>
<tr><th>Level</th><th>Masculine narrative</th><th>Feminine narrative</th></tr>
<tr><td>End 3-22</td><td>


</td><td>


</td></tr>
<tr><td>End 3-29</td><td>


</td><td>


</td></tr>
</table>

Level	Masculine narrative	Feminine narrative
End 3-29		No second image for feminine narrative
End 3-29		

APPENDIX B:
RUBRIC FOR OPEN-ENDED NARRATIVE QUESTION

Rubric: NQ 9, open ended
Please explain why you wanted to reach the end of the game. Or explain why you didn't care about getting to the end of the game.

Variable	Explanation	Score Yes	Score No	Cronbach's kappa
Narrative_endgame_pos	Any positive explanation. Example: • **I did because I wanted to complete it** but I didn't because I didn't get really hooked on it.	1	0	.956
Narrative_endgame_neg	Any negative explanation. Examples: • I really didn't care about this game. • Because it was boring. • I did because I wanted to complete it but **I didn't because I didn't get really hooked on it.** • I thought that Patch was cute but **the game got me confused sometime. I didn't really get it.**	1	0	.830
Narrative_character_pos	Positive opinion of character in their explanation. Example: • **I thought that Patch was cute** but the game got me confused sometime. I didn't really get it.	1	0	.886
Narrative_character_neg	Negative opinion of character in their explanation. Examples: • Patch is a girl, I wanted the boy. • Patch was annoying.	1	0	1.000
Narrative_story_pos	Positive opinion of narrative in their explanation. Examples: • I thought the story was interesting. • I wanted to find out what was going to happen to Patch.	1	0	.910
Narrative_story_neg	Negative opinion of narrative in their explanation. Example: • I didn't care what happened to Patch.	1	0	.936
Math_pos	Positive opinions of math. Example: • I did because I like math!	1	0	1.000
Math_neg	Negative opinions of math. Example: • I don't care because I hate math.	1	0	.906

Completely blank for question, give 99s across the board.

APPENDIX C:
BACKGROUND INFORMATION SURVEY

1. Birth date: ________ / ________
   Month Year

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

3. What are you learning in your math class now? ______________________________________

4. Gender: ☐ Male ☐ Female

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

6. 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

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

☐ A ☐ B ☐ C ☐ D ☐ F ☐ Don't know

8. What were your math grades last year?

☐ A ☐ B ☐ C ☐ D ☐ F ☐ Don't know

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

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

APPENDIX D:
VIDEO GAME EXPERIENCE SURVEY

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

   ☐ Poor ☐ Fair ☐ Good ☐ Very good

2. 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

3. How many **HOURS a WEEK** do you play video games on the following:

Platform	**0 hours/week**	**1-4 hours/week**	**5-8 hours/week**	**9-12 hours/week**	**13 or more hours/week**
…Nintendo	1	2	3	4	5
…Playstation	1	2	3	4	5
…Xbox	1	2	3	4	5
…a computer	1	2	3	4	5
…PSP	1	2	3	4	5
…DS	1	2	3	4	5
…iPhone/iPod	1	2	3	4	5
…any **other** handheld or mobile device	1	2	3	4	5

4. What video game do you play the most **right now** on the following:

…Nintendo	______________________
…Playstation	______________________
…Xbox	______________________
…a computer	______________________
…PSP	______________________
…DS	______________________
…iPhone/iPod	______________________
…any **other** handheld or mobile device	______________________

5. What is your **favorite video game of all time**? ____

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

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

☐ Poor ☐ Fair ☐ Good ☐ Very good

<u>Questions 22-24 are only about VIDEO AND COMPUTER GAMES:</u>

8. How often do you play puzzle games? (*examples*: Tetris, Minesweeper, Bejeweled)

☐ Never ☐ Hardly ever ☐ Sometimes ☐ Often ☐ Very often

9. How often do you play math games at school?

☐ Never ☐ Hardly ever ☐ Sometimes ☐ Often ☐ Very often

10. How often do you play math games at home?

☐ Never ☐ Hardly ever ☐ Sometimes ☐ Often ☐ Very often

APPENDIX E:
STUDY MISSING DATA ANALYSES

Missing data. An examination of students' missing data responses for the math pretest and posttest items was conducted. Missing was defined as a blank response where there was no indication that the student knew or did not know the answer to the item. Eighty-two (46.3%) participants had one or more missing responses on the pretest, 71 (40.1% of sample) participants had one or more missing responses on the posttest (with respect to the same items on the pretest), and 34 (19.2% of sample) participants had one or more missing responses on posttest items that were administered only on the posttest. Complete data (i.e., no missing items on any of the measures) were available for 71 participants (40.1% of sample). These high numbers of missing items for the pre- and posttests are not surprising as many of the students were in remedial math courses and struggled to complete every question in the amount of time given. The number of missing responses for the posttest only items was relatively low, justifying the validity of our choice to use these items as our main outcome measure.

We inspected the distribution of missing responses on the pretest and posttest as shown in Table E1. We chose a missing response rate of 20% for each measure (i.e., allowing at most five missing items for pre- and posttest, and two items for the posttest only items), as there appeared to be a drop in the number of missing responses after five. This threshold resulted in dropping 34 participants and allowed us to retain 81% of the original sample.

We then checked whether there were differences between the original sample and the reduced sample with respect to distribution across conditions, and distribution of grades, gender, and gameplay experience across conditions. We also checked if the sample characteristics (means and standard deviations) changed significantly before and after dropping cases. Table E1 shows the distribution of the number of missing responses by measure.

Table E1

Number of Missing Responses in Pretest, Posttest, and Posttest Only Items (N = 177)

		No. of missing responses							
Scale	No. of items	0	1	2	3	4	5	6	>6
Pretest	23	95	42	13	7	5	5	1	9
Posttest	23	106	28	13	8	3	2	4	13
Posttest only	10	143	9	3	3	9	1	4	4

Table E2 shows the distribution of the number of missing responses by condition. A chi-square test for the difference in the distribution of the different samples across conditions was not significant (χ^2 = 4.15 p = .13), suggesting that the distribution of participants with missing data was not systematically different (by condition) from the distribution of participants with complete data.

Table E2

Distribution of Missing and Retained Participants by Condition

	Condition			
Sample	Masculine	Feminine	No narrative	Total
Missing sample	17	10	7	34
Remaining sample	40	45	58	143
Total	47	62	68	177

Table E3 shows the means and standard deviations of the reduced sample and the missing sample. Independent t tests were conducted to check to see if there were differences between the samples on the math measures. There were significant differences between the retained sample and the missing sample on both measures. These differences are unsurprising, as the missing data are treated as incorrect (i.e., biasing the score lower).

Table E3

Comparison of Reduced and Missing Samples: Means and Standard Deviations

	Sample without missing data			Missing sample		
Measure	*n*	*M*	*SD*	*n*	*M*	*SD*
Pretest	143	11.05	5.01	34	7.79	4.72
Posttest	143	12.10	5.19	33	7.62	4.08
Posttest (items only on posttest)	143	5.32	2.65	33	2.71	2.49

Table E4 shows the means and standard deviations of the reduced sample and the original sample. Independent *t* tests were conducted to check to see if there were differences between the samples on the math measures. No significant differences were found on any measure. Variance ratio tests were also conducted to check for differences in variances between samples; no significant differences were found. Thus, the removal of missing cases did not appear to affect the mean or variance of the sample.

Table E4

Comparison of Reduced and Original Samples: Means and Standard Deviations

	Sample without missing data			Original sample		
Measure	*n*	*M*	*SD*	*n*	*M*	*SD*
Pretest	143	11.05	5.01	177	10.43	5.11
Posttest	143	12.10	5.19	176	11.26	5.29
Posttest only	143	5.32	2.65	176	4.83	2.80

Table E5 shows the skewness and kurtosis of the two samples. Fisher's skewness coefficient and Fisher's kurtosis coefficient were computed to test for differences between the original and reduced samples. There was no significant skewness for the original or reduced sample, suggesting the removal of missing cases did not affect the skewness of the distribution. In terms of kurtosis, both samples appeared platykuric, suggesting both samples deviated from the normal distribution. Based on these analyses, we concluded that the removal of cases due to missing values on the math measure did not substantially change the shape of the distribution of the reduced sample.

Table E5

Comparison of Samples: Skewness and Kurtosis

	Original sample					Sample without missing data				
Measure	*n*	Skewness	*SE*	Kurtosis	*SE*	*n*	Skewness	*SE*	Kurtosis	*SE*
Pretest	177	.45	.18	-.83	.36	143	.44	.20	-.93	.40
Posttest	176	.50	.18	-.66	.36	143	.46	.20	-.86	.40
Posttest only	176	.22	.18	-1.11	.36	143	.08	.20	-1.22	.40

The next set of analyses examined whether dropping the cases with missing data resulted in different distributions across conditions within subgroups with respect to gender, game experience, and self-reported grades in math. As Table E6 shows from results of one-way ANOVAs, the samples were similar with respect to the distribution of participants by weekly gameplay, self-reported grades, and gender regardless of whether they received a masculine narrative, feminine narrative, or no narrative.

Table E6

Comparison of Samples: Condition by Gender, Weekly Gameplay, and Self-Reported Grades in Math

	Original sample				Sample without missing data			
Condition	*n*	Statistic	*df*	*z*	*n*	Statistic	*df*	*p*
Condition × gender	165	F = .221	2	.80	134	F = .38	2	.69
Condition × self-reported math grades	157	F = 1.47	2	.23	126	F = .87	2	.42
Condition × weekly gameplay frequency	167	F = 5.36	2	.59	136	F = .12	2	.89

From these analyses we concluded that dropping 15 cases from the original data set because of missing responses on the math measures did not unduly affect the mean, variance, the shape of the distribution on the math measures, or the distribution of participants across conditions and various subgroups. Thus, subsequent analyses were based on data from 143 participants.