

## Abstract Title Page

**Title:**

***Battleship Numberline: A Digital Game for Improving Estimation Accuracy on Fraction Number Lines***

**Author(s):**

**Derek Lomas, Dixie Ching, Eliane Stampfer, Melanie Sandoval, Ken Koedinger**

## Abstract Body

### **Background / Context:**

In 2008, the National Mathematics Advisory Panel stated: “the most important foundational skill not presently developed appears to be proficiency with fractions.” In response to numerous studies describing the challenges faced by American students in fractions learning, in 2010 the Institute for Education Sciences (IES) released a practice guide for “*Developing Effective Fractions Instruction for Kindergarten through 8th Grade*.” This practice guide strongly advocates the use of number lines for improving students’ understanding of fractions (Siegler et al., 2010). Teachers in America tend to use part-whole representations of fractions (e.g., pizza slices) as opposed to number lines, which are a more common instructional tool in Asian countries (Ma, 1999; Moseley, Okamoto, & Ishida, 2007; Watanabe, 2006). Number lines are also used as assessment tools for measuring number sense. Notably, recent work by Siegler, Thompson, and Schneider (in press) shows that the accuracy of number line estimation with fractions correlates with standardized test scores in 6-8<sup>th</sup> grade. This finding extends prior research on number line estimation with decimals (Schneider, Grabner, & Paetsch, 2009) and whole numbers (Booth & Siegler, 2008), which found that accuracy predicted standardized test scores in grades K-5.

Given the strong relationship between number line estimation accuracy and math achievement, might a computer-based number line game help improve math achievement? In one study by Rittle-Johnson, Siegler and Alibali (2001), a simple digital game called *Catch the Monster* provided practice in estimating the location of decimals on a number line. The authors found that improving number line estimation accuracy transferred to other conceptual skills, such as the ability to compare the magnitudes of different decimal values. To our knowledge, however, no research has yet demonstrated any benefits from improving number line estimation accuracy in the domain of fractions.

As the first step in an ongoing research study to investigate the benefits of improving number line estimation skills with fractions, this paper presents preliminary findings gathered by a novel digital game, *Battleship Numberline* (see Figure 1). This study advances the area of number line research by demonstrating that students can make significant gains in estimation accuracy with just 20 minutes of game play.

### **Purpose / Objective / Research Question / Focus of Study:**

In this pilot study, we seek to investigate whether students can successfully improve their fraction estimation accuracy within *Battleship Numberline*. Additionally, we investigate whether the data generated by the game provides a meaningful assessment of a student’s understanding of fractions.

### **Setting:**

This experiment was conducted in a Pittsburgh charter school using an online version of *Battleship Numberline*, accessed through a web browser. Students played *Battleship Numberline* for up to 20 minutes during their math class, either on laptop computers in their classrooms or desktops in a school computer lab.

**Population / Participants / Subjects:**

A total of 51 students ( $M=10.5$  years,  $SD=1.5$ ) were selected by their teachers to participate in this study. A mix of ages and abilities were included, for a total of 20 fourth graders, 19 fifth graders, and 12 sixth graders.

**Intervention / Program / Practice:**

In sessions lasting less than 20 minutes, each student encountered 86 trials of number line estimation practice during four levels of *Battleship Numberline*. The first six trials served as an introduction to the game mechanic: students were presented with a number line from 0-10 and had to estimate the location of six different whole numbers. In each of the subsequent four levels of fraction estimation, the same 20 fractions were presented in a random order:  $1/10$ ,  $1/8$ ,  $1/6$ ,  $1/5$ ,  $1/4$ ,  $3/10$ ,  $1/3$ ,  $3/8$ ,  $2/5$ ,  $3/7$ ,  $1/2$ ,  $3/5$ ,  $5/8$ ,  $2/3$ ,  $7/10$ ,  $3/4$ ,  $4/5$ ,  $5/6$ ,  $7/8$ ,  $9/10$ . Each level was either a typing (T) level or a clicking (C) level (see Figure 1). In a typing level, students had to type their answer in response to the position of a visible ship on the number line, while in a clicking level students had to click on the location of a hidden submarine, in response to the presented fraction.

After the student clicks or types their estimate, a bomb drops at the indicated location. Various kinds of feedback are then presented to the student, including animations of the hits and misses, the percent accuracy of their current estimate, a running total of their average accuracy, and corrective numerical feedback on the fractional location of the ship.

**Research Design:**

Each student was randomly assigned to one of four instructional sequences, which systematically varied the presentation order of the typing and clicking levels (TTCC, CCTT, TCTC, and CTCT). Therefore, each student encountered the same 80 trials of fraction estimation practice, but in a different order. Using a within-subjects design, we can test the following hypotheses:

*Hypothesis 1:* 20 minutes of game play produces measurable learning

*Hypothesis 2:* Typing is harder than clicking

*Hypothesis 3:* Unit fractions are easier than non-unit fractions

**Data Collection and Analysis:**

*Battleship Numberline* is a Flash game played online in a web browser. Student performance data is collected in a MySQL database. Data was anonymized and imported into the Pittsburgh Science of Learning Center DataShop (<http://pslclatashop.org>), where it may be accessed and analyzed by other researchers.

On each trial, we collected whether a response was successful or not (hit rate), the accuracy of the response, and the student's reaction time. Each ship is centered on the location of the tested item (e.g.,  $1/3$ ) and the length of the ship is set to be 10% of the length of the number line. Therefore, a student will receive a successful "hit" if their estimate has no more than 5% Percent Absolute Error (PAE). PAE is calculated by the absolute value of the estimated location minus the actual location, divided by the difference between the number line endpoints; accuracy is the PAE subtracted from 100. For example, on a number line from 0-1, if a ship is at  $1/3$  and a player clicks at  $1/4$ , the PAE is 8% (a miss) and the accuracy is 92%. If a student runs out of time after 15 seconds, they are given 66% accuracy for the trial (the accuracy of chance performance).

In order to calculate learning gains without the use of a pre/post test, we averaged together a student's performance on their first typing level and their first clicking level and named this "First Opportunity." Then, performance data on the second typing level and the second clicking level were averaged and named "Second Opportunity." By comparing performance gains between the First Opportunity and the Second Opportunity, we obtained a measure of their improvement *within* the game.

Additionally, a paper-based self-report questionnaire was provided to students after they played in order to gain a rough measure of students' affective response to the game. Example items include measures of enjoyment and replayability, such as "This game was fun to do" and "I want to play this game again some time." These and 19 other items were answered on a 1-5 scale, labeled Strongly Disagree, Disagree, I'm in the middle, Agree, and Strongly Agree.

### **Findings / Results:**

In Table 1, we document the significant improvements in number line estimation accuracy and hit rate that occurred during 20 minutes of *Battleship Numberline* game play. In most grade levels, we found significant improvements in both accuracy and hit rate. The year-to-year difference in accuracy between the grade levels is an average of 9.2%; therefore, we find it notable that the 4<sup>th</sup> graders rose by 7.3% or about four-fifths of a grade level of performance during the 20 minutes of play. When considering hit rate improvements, the 4<sup>th</sup> graders' 6.8% improvement is more than a third of the average 18.2% difference in hit rates between grade levels. To investigate whether these effects are lasting and transfer, we will have students play the game again after three months and give a paper-based test. We will present this data at the conference.

In addition to considering gross improvements by students, we can "zoom in" to our data to investigate the specific nature of student difficulties. Figure 2 shows hit rate for each fraction on their first and second opportunities and reveals an improvement for most fractions. As one would expect, students are more accurate when estimating  $1/2$  than  $3/7$ . Perhaps less obviously, we found that students achieve higher hit rates on items nearer to the end points. Apparently, students have some good intuitions for adjusting a fraction down from one (e.g.,  $9/10$  is just below 1), but *not* for adjusting from  $1/2$  (e.g.,  $3/7$  is just below  $1/2$ ).

As shown in Table 2 we found the hit rate on unit fractions (e.g.,  $1/3$ ,  $1/5$ ), which is 50%, significantly higher than the hit rate on non-unit fractions (e.g.,  $2/3$ ,  $2/5$ ), which is 42%. Accuracy reveals the same pattern. This difference may be indicative of needing to first learn the function of the denominator (e.g., how many divisions) and then of the numerator (e.g., how many divisions to count). Table 2 also shows performance comparisons on clicking vs. typing levels. Curiously, while hit rate performance is worse on clicking than typing, accuracy performance is the reverse with typing being harder than clicking. We expected typing to be harder (as it is for accuracy) because production is usually more difficult than comprehension. However, the hit rate measure may have favored typing in that there are fewer discrete choices when producing a typical fraction (i.e., not so many numerators and denominators to choose from) whereas there is a more continuous set of choices in clicking. In other words, there are more ways to be a little off in clicking than when typing.

We have begun to investigate patterns of error, which may be evidence of learner misconceptions. As many children have the misconception that increasing the size of the denominator in a fraction increases its value, we expected that some portion of the subjects would view  $1/10$  as a large fraction. Indeed, when investigating the click responses to  $1/10$ , we found that the region of the line around  $9/10$  received the most incorrect clicks. A similar pattern was revealed for the fraction  $1/8$ .

While a full analysis of the self-report questionnaire cannot be presented here, we report summary statistics in Table 3. Perhaps not surprisingly boys tended to rate the game more favorably than girls. Nevertheless, about  $2/3$  of the girls agreed the game is good/fun and about  $3/4$  agreed they would play it again.

### **Conclusions:**

During our pilot study, *Battleship Numberline* appeared to produce academically relevant improvements in learning during a relatively short period of time. Additionally, the game generated data that may be useful as a cognitively diagnostic assessment, such as for identifying common misconceptions about fractions. This data also illustrates general patterns of difficulty over different types of estimation tasks, whether estimating unit or non-unit fractions or estimating fractions through clicking or typing. Finally, we have presented limited evidence that our game is, in fact, enjoyable for students to play. One limitation of our current design is that we can only confirm that learning is occurring within the game. In the next study we will use a paper pre/post test to determine whether learning successfully transfers to measures of fraction understanding outside the game.

By creating a game-based intervention, we have the goal of making large amounts of numerical practice playful and palatable, both in and out of school. There is a large amount of interest in learning games as learning objects (Mayo, 2009), but very little data to empirically support their effectiveness or efficiency. While significantly more study is needed to demonstrate the longevity or utility of the learning produced by *Battleship Numberline*, this study demonstrates the rich data available from microgenetically tracking learning within a game. This data could be used for large-scale online research studies, or for presenting ongoing assessments of learner performance to teachers, parents and even to learners themselves.

This pilot study implicates a broader set of learning game design research. During the design process of *Battleship Numberline*, we identified numerous game design features that, when parametrically implemented, could affect student learning and engagement outcomes. Given online use at a reasonable scale ( $>10,000$  users), different design features could be evaluated for their effects on learning and engagement. Different experiments could then be structured to investigate research questions in the educational sciences or simply used as a practical process for the continuous improvement of the game over time.

## Appendices

*Not included in page count.*

### Appendix A. References

*References are to be in APA version 6 format.*

- Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*, 1016-1031.
- Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Mayo, M. J. (2009) Video games: a route to large-scale STEM education? *Science, 323*, 5910, 79-82.
- Moseley, B. J., Okamoto, Y., & Ishida, J. (2007). Comparing US and Japanese elementary school teachers' facility for linking rational number representations. *International Journal of Science and Mathematics Education, 5*, 165-185.
- National Mathematics Advisory Panel (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: National Council of Teachers of Mathematics.
- Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*, 346-362.
- Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in Grades 5 and 6. *Journal of Educational Psychology, 101*, 359-372.
- Siegler, R., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., ... Wray, J. (2010). *Developing effective fractions instruction for kindergarten through 8th grade: A practice guide* (NCEE #2010-4039). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/wwc/publications/practiceguides>
- Siegler, R. S., Thompson, C. A., & Schneider, M. (in press). An integrated theory of whole number and fractions development. *Cognitive Psychology*.
- Watanabe, T. (2006). The teaching and learning of fractions: A Japanese perspective. *Teaching Children Mathematics, 12*, 368-374.

## Appendix B. Tables and Figures

Not included in page count.

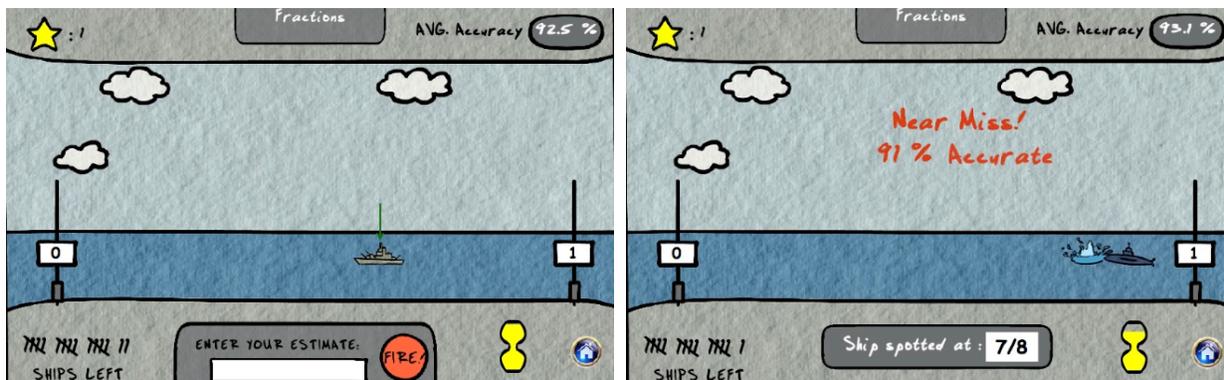


Figure 1: The image on the left shows the typing level, where a player will enter their estimate into the box in response to the position of the ship on the number line. The image on the right shows the clicking level after a player clicked to estimate the location of the presented fraction, 7/8. Prior to their click, the submarine was hidden.

Table 1: The estimation performance differences over the first and second opportunity, according to grade level.

Grade	N	Average Accuracy			Paired t-test	Average Hit Rate			Paired t-test
		First Opportunity	Second Opportunity	Change		First Opportunity	Second Opportunity	Change	
4	20	73.3% (.13)	80.6% (.12)	5.3%	p<.01	27.6% (.18)	34.4% (.19)	6.8%	p<.02
5	19	83.9% (.13)	86.3% (.13)	2.5%	n.s. p=.053	43.6% (.24)	51.8% (.22)	8.3%	p<.01
6	12	91.8% (.04)	91.8% (.01)	2.0%	p<.02	64.0% (.10)	69.0% (.12)	5.0%	n.s. p=.18
<b>Total</b>	51	81.6% (.13)	81.6% (.12)	3.5%	p<.0005	42.1% (.23)	49.0% (.23)	6.9%	p<.001

Average Item Hit Rate from First to Second Opportunity

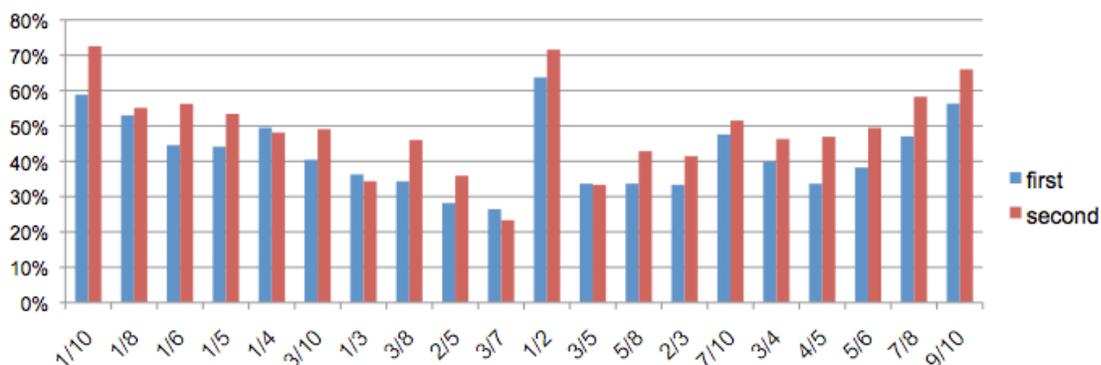


Figure 2: The change in performance on individual fractions over the first and second opportunity.

Table 2: The significant performance differences on unit vs. non-unit fractions and clicking vs. typing.

Fraction Type	Average Accuracy	Average Hit Rate	Level Type	Average Accuracy	Average Hit Rate
Unit	85.0% (.13)	50.5% (.21)	Clicking	86.3% (.09)	39.9% (.23)
Non-Unit	83.2% (.12)	41.6% (.26)	Typing	81.1% (.18)	49.0% (.26)
Paired t-test	p<.02	p<.0001	Paired t-test	p<.02	p<.001

**Table 3: Percent of students responding “Agree” or “Strongly Agree” to listed self-report items**

<b>Questionnaire Item</b>	<b>All (n=51)</b>	<b>Female (n=27)</b>	<b>Male (n=24)</b>
"This was a good game"	78%	67%	92%
"This game was fun to do"	74%	63%	87%
"I think my friends would like this game"	49%	41%	58%
"I want to play this game again some time"	82%	74%	92%