Fraction Ball: Playful and Physically Active Fraction and Decimal Learning

Andres S. Bustamante
Kreshnik Nasi Begolli
Daniela Alvarez-Vargas
Drew H. Bailey
Lindsey Engle Richland

University of California, Irvine
School of Education

Accepted at *Journal of Educational Psychology* 08/19/2021

Author Note

This work was supported by the UCI Council on Research, Computing, & Libraries and the UCI School of Education Dean's Research Fund. Authors ASB & KB contributed equally to this manuscript. All authors contributed to the design and implementation of the study, as well as analyses and preparation of the manuscript.

Correspondence in regards to this manuscript should be addressed to asbustam@uci.edu

Acknowledgments

We would like to thank the teachers, students, and administrators at El Sol Academy for their partnership, creativity, and enthusiasm. They made this work possible. Specifically, we acknowledge Lourdes Farag, Guadalupe Cruz, Ivet Gonzalez, Crystal Quinonez, Patricia Martinez, Susan Saldana, Shamin Huq, Jenny Zavala, Tristan Gude, and Monique Daviss for their great contributions to this work. We also thank Lisa Fazio and Hugues Lortie-Forgues for their feedback during the early stages of this work. We also want to acknowledge the Advanced Education Research and Development Fund (AERDF) who is supporting the next phase of this work through their EF+Math Program to bring Fraction Ball to the Santa Ana Unified School District.

Abstract

This study tested a novel approach to capitalizing on the benefits of play for informal math learning. Two experiments evaluated a platform called “Fraction Ball”, that provides an embodied, playful, and physically active learning experience by modifying the lines on a basketball court to support rational number learning. In the Pilot Experiment, 69 5^{th}-6^{th} graders were randomly assigned to play a set of 4 different Fraction Ball games or attend normal physical education (PE) class and completed rational number pretests and posttests. After strategic improvements to expand the intervention, the same protocol was implemented in the Efficacy Experiment with 160 4^{th}-6^{th} graders. Playing Fraction Ball for 4 PE class periods (Pilot Experiment) improved students’ ability to convert fractions to decimals. Playing a revised version of 6 different Fraction Ball games for 6 PE class periods (Efficacy Experiment) significantly improved children’s rational number understanding as reflected by higher scores in overall accuracy, with positive impacts on several subtests. Fraction Ball represents a low-cost, highly scalable intervention that promotes math learning in a fun and engaging approach.

Keywords: Rational Numbers, Playful Learning, Randomized Control Trials, Math Cognition

Educational Impact and Implications Statement

Fraction and decimal number learning are notoriously difficult for elementary aged students. Fraction Ball represents a new way to engage children in fraction and decimal number learning that is playful, engaging, embodied, physically active, and rooted in cognitive science research. This study not only provides evidence for the efficacy of Fraction Ball in promoting rational number learning through a rigorous experimental design, but also demonstrates the potential of the school yard as a context for play-based interventions that can promote learning across academic domains using an engaging and evidence-based approach. Further, this study serves as a model for how interventionists can engage educators as partners in intervention design and iterative implementation to improve learning outcomes for students. Fraction Ball is affordable, easy to disseminate, and has the potential to promote rational number learning on a global scale.

Fraction Ball: Playful and Physically Active Fraction and Decimal Learning

Children learn effectively in active, meaningful, engaged, and socially interactive settings—making play an ideal context to foster learning and development (Hirsh-Pasek et al., 2015). Integrating fundamental mathematical topics into playful learning contexts presents new opportunities for improving education while capitalizing on children's drive to play and engage physically with their environment (Bustamante et al., 2020; Hassinger-Das et al., 2018; Schlesinger et al., 2020).

This study introduces and attempts to measure the effectiveness of playing a set of Fraction Ball games, a novel intervention that provides a playful and physically active learning experience—rooted in math cognition research—by modifying the lines on a basketball court to emphasize rational number learning (see Figure 1). Findings are educationally significant because 1) rational numbers are challenging but of foundational importance to long-term mathematics understanding, 2) they inform the rationale for and design of play-based mathematics interventions at the elementary level, and 3) they provide novel theoretical insights into the contributors to rational number understanding.

Educational need

Children's difficulties with rational numbers in the forms of fractions and decimals are pervasive (Siegler et al., 2012;), inequitable, (Sonnenschein & Galindo, 2015), and function as a gatekeeper for understanding more complex mathematics like algebra (Booth & Newton, 2012) and advanced science concepts (NGSS, 2013). In general, many U.S. children have a limited conceptual understanding of fractions and decimals. For example, less than half of 8th graders correctly ordered the magnitudes of three fractions in the National Assessment of Educational Progress assessment (NCTM, 2006). When asked whether .274 or .83 was larger, most 5th and

6th graders choose .274 (Lortie-Forgues et al., 2015; Rittle-Johnson et al., 2001). The difficulty extends to adolescents and adults; in a study of U.S. 11th graders, fewer than 30% translated .029 into the correct fraction (Klostermann, 2010), and these patterns were observed in U.S. community college students as well (Givvin et al., 2011). This suggests new strategies for supporting rational number concepts are needed in schools.

Research Foundations for Fraction Ball

Fraction Ball draws on well-established literatures to address these rational number concept skills. We review them briefly.

Fraction Magnitude Understanding

The Integrated Theory of Numerical Development (Siegler et al., 2011; Siegler & Lortie-Forgues, 2014) posits that mental number line representation is foundational for rational number learning, and interventions improving children's understanding of rational number magnitudes can raise performance on a variety of fraction and decimal tasks, many of which are required for performance in more advanced mathematics (Siegler et al., 2011, 2012). In the case of fractions, key concepts include understanding that fractions represent a unit divided into equally sized parts and can be represented by different fraction symbols. Students must learn how the numerator and denominator are related to each other and to fraction magnitudes in order to place fractions in order on a number line (Hansen et al., 2015; Siegler et al., 2011). Consensus in the field suggests that these fraction concepts may be a foundation on which children can build a wide range of conceptual and procedural knowledge of fractions (Seethaler et al., 2011; Siegler et al., 2011; Vamvakoussi & Vosniadou, 2010). Yet, many studies suggest that middle-school-aged children and even adults training to be teachers struggle with conceptual understanding of fraction magnitudes and fraction arithmetic operations (Siegler & Lortie-Forgues, 2015; Siegler & Lortie-

Forgues, 2017). Many have argued this is largely a result of a culture of teaching and learning fractions (and mathematics more broadly) as disconnected pieces of information (Givvin et al., 2011; Richland et al., 2012).

Common Misconceptions with Fractions and Decimals

Rational number operations are often taught consisting of procedural steps that need to be memorized and then followed in order to get the correct answer (Richland et al., 2012). Unique procedures are taught for solving problems represented in fraction and decimal notations and the systematic errors that children make with each notation can at times be traced to a mathematical source common to both, yet other systematic errors are unique to each notation (Tian & Siegler, 2017). Children learning the magnitudes of individual fractions and decimals commonly over-generalize properties of whole numbers (e.g., larger whole numbers magnitudes always indicate that a number has a greater magnitude) a phenomenon sometimes called “whole number bias” (Ni & Zhou, 2005). For example, a common misconception is that a/8 is larger than a/5, presumably because 8 is larger than 5 with about 47% of community college students choosing a/8 with 50% being chance (Stigler et al., 2010). Similarly, 0.50 can be thought to be smaller than 0.5 in decimal format, because 50 is larger than 5, with data suggesting that only 43% of 9^{th} graders and about 50% of Australian pre-service teachers were able to correctly order decimals with a different number of digits after the decimal (e.g., 0.034, 0.485, 0.1423, 0.4 or 0.6606, 0.606, 0.66, 0.6; Hiebert & Wearne, 1985; Tian & Siegler, 2017).

Further, frequent errors in students’ fraction arithmetic procedures reflect a lack of conceptual understanding of the relations between fraction magnitudes and fraction arithmetic procedures (Lortie-Forgues et al., 2015; Siegler & Lortie-Forgues, 2014). A common incorrect strategy used by students learning to add and subtract fractions is to add both the numerators and denominators (e.g. $\frac{1}{2} + \frac{1}{4} = \frac{2}{6}$). One method proposed for preempting this error is to encourage students to think of fractions as sets of unit fractions; for example, $\frac{3}{4}$ is 3 of the $\frac{1}{4}$ unit. Successful interventions for improving fraction arithmetic have incorporated this practice (Braithwaite & Siegler, 2020; Fuchs et al., 2016). With decimal notation, whole number bias manifests in a different form, such that when the number of decimal places differs, (e.g., 4.2 + 3.42) only 12% of 6^{th} graders could solve the problem correctly, but 74% were correct when the number of decimal places was the same (e.g., 3.44 + 2.21; Hiebert & Wearne, 1985).

Children sometimes appear to treat fractions and decimals as belonging to separate number systems (Fuchs et al., 2016). For example, when asked "how many numbers are in a given interval," most of a sample of 8^{th} graders responded that only fractions can be placed between an interval of two fractions (e.g., 1/3 – 2/3) and only decimals can be placed between an interval of two decimals (e.g., 0.1 – 0.2; Vamvakoussi & Vosniadou, 2010). Perhaps unsurprisingly, children also have difficulties converting between the two symbolic systems despite them representing the same magnitude (Hiebert & Wearne, 1983). Many have argued that understanding the relationship between fractions and decimals may serve to highlight the commonalities between the two notations to reinforce magnitude understanding as well as detect errors. For example, checking answers in a more familiar notation may set expectations for plausible upper and lower bounds of potential answers in the less familiar notation (Wang & Siegler, 2013).

Comparing Multiple Representations

The cognitive principle of using analogies and/or comparing and contrasting cases has been advocated by mathematics experts as a powerful tool to improve mathematics outcomes (Booth et al., 2018; CCSSM, 2010; NMAP, 2008; NRC, 2001; Richland & Begolli, 2016) with particular implications for rational number representation (DeWolf et al., 2015), schema formation (Begolli & Richland, 2016), and for overcoming misconceptions (Vamvakoussi, 2017). The power in successfully engaging in comparing and contrasting representations is believed to direct the reasoner to notice commonalities and differences which highlights irrelevant (or surface) features and relevant (or structural) features of a concept (Richland & Begolli, 2016). In the case of fraction and decimal notations, children may notice that that ¼ and 0.25 are written differently (surface features), but they represent the same magnitude on a mental number line (structural commonality) (Wang & Siegler, 2013). A notable study by Moss and Case (1999) used a curriculum that emphasized the connections between fractions and decimals (as well as percentages). For example, the curriculum used fraction and decimal terms interchangeably (e.g., one-fourth ¼ was used when 0.25 was presented) and visuals were labeled with all rational number notations (Moss & Case, 1999). This led children in the experimental group to show a greater understanding of multiple rational knowledge measures and was replicated with a different sample (Kalchman et al., 2001). However, successfully translating the principle of comparing and contrasting into everyday teaching and curricula has been challenging in the U.S. (Richland et al., 2012; Star et al., 2015). The number line could serve as a model for drawing connections between fraction and decimal symbolic systems since these notation comparisons can also highlight the common magnitudes these notation systems represent (Wang & Siegler, 2013).

Number Line Interventions

The number line has proven to be an effective tool for teaching children about whole number magnitudes (Link et al., 2013; Siegler & Ramani, 2008) and fraction magnitudes (Fazio et al., 2016) in the context of games. For example, Fazio and colleagues (2016) observed transfer from computerized fraction number line training to other measures of children's fraction magnitude knowledge. Intensive intervention focused on fraction magnitude knowledge can influence both fraction magnitude and fraction arithmetic knowledge (Barbieri et al., 2020; Fuchs et al., 2013; 2017; Malone et al., 2019; Schumacher et al., 2018). These interventions have been beneficial for 4^{th}-6^{th} graders who struggled with fraction concepts (Barbieri et al., 2020; Dyson et al., 2020; Fuchs et al., 2013; 2017) as well as students with disabilities in grades 6^{th}-8^{th} (Bottge et al., 2014). While the above studies vary in many other components where each study integrates differing and sometimes overlapping principles stemming from the science of learning (e.g., self-explanation prompts; Barbieri et al., 2020) the visual representation of the number line is common to all of these interventions that showed improvements on rational number skills. Further, the number line representation seems to supersede other rational number representations (e.g., circular area models); for example, data from Hamdan and Gunderson (2017) suggest that number line training resulted in greater transfer to novel fraction comparison tasks than area model training with second and third graders. As such, the number line representation seems a crucial component in intervention design that can serve to integrate multiple factors necessary for learning about rational numbers.

Embodied Cognition

Theories of grounded or embodied cognition suggest that sensory and bodily experiences underpin more abstract cognitive concepts like numerical magnitudes (Barsalou, 2010; Case & Okamoto, 1996; Dehaene, 1997; Wilson, 2009). Studies of embodied cognition in the context of numerical cognition describe the potential advantages of including correlated "kinesthetic, auditory, visuo-spatial, and temporal cues" that all contribute to a multimodal learning experience (Siegler & Ramani, 2009, p.6). Some previous interventions have engaged children in whole-body participation in number-line spatial training to improve children's numerical representations. For example, Link and colleagues (2013) compared whole number line training where one group walked a life-size number line while the other group experienced the same training without the full body experience. Results demonstrated a significant effect in favor of the embodied number line condition. In two separate studies, attempts to combine the external bodily experience of moving along the number line with an individual's internal representation of number magnitudes has been shown to help students improve children's number line estimation (Fischer et al., 2011) and addition (Siegler & Ramani, 2009). Interestingly, the embodied cues alone do not seem to be sufficient for improving learning, rather these cues may require an essential component, such as the linearity of the number line and data from Ramani & Siegler (2008; 2009) support this possibility where circular games were not beneficial for learning compared to linear games. Despite these promising lines of research, there is limited work in embodied fraction number line games.

Fraction Ball Intervention

The Fraction Ball design was guided by synthesizing several different areas of research including the importance of number lines for fraction learning (Fazio, 2016), common misconceptions with rational numbers (Durkin & Rittle-Johnson, 2015; Lortie-Forgues et al., 2015), the feasibility and effectiveness of whole number line training and games (Link et al., 2013), the potential of learning from comparisons (Richland & Begolli, 2016), and the power of embodied cognition and playful learning spaces (Bustamante et al., 2020; Glenber, 2010). Generally, games are a potent tool for engaging children in learning (Hassinger-Das et al., 2017) and play increases children's motivation to participate in learning situations, particularly for children with attention or behavior problems who often struggle with more traditional classroom settings (Sarihi et al., 2015; Zosh et al., 2018). Game features were also specifically designed to capitalize on play to ensure students' engagement with key rational number concepts.

The spatial layout of the court and the point system of the game were designed to reinforce magnitude understanding and fluency with arithmetic based on the Integrated Theory of Numerical Development (Siegler et al., 2011). All point values were represented in fraction and decimal form, thus, allowing the players of the game to represent and add points intuitively in fraction and decimal format. The court layout was such that as children moved further from the basket, the fraction and decimal magnitudes increased in line with the spatial distance from the basket (Link et al., 2013). The fraction denominators were fourths on one side of the court and thirds on the other side of the court to promote engagement with two different fraction and decimal representations. Each side of the court was split in half, such that one side was labeled with fractions and the other side was labeled with decimals, thus, providing an opportunity to compare and contrast the symbolic representations of the equivalence between fractions and decimals with respect to the magnitudes they represent (Richland & Begolli, 2016). These concepts of fraction and decimal magnitudes were further reinforced in the game scoreboard, which was a walkable number line with hash marks that divide whole numbers in fourths (or thirds) with fraction and decimal notations next to each hash mark (Figure 1).

Current Study

This study evaluated the learning utility of playing a set of Fraction Ball games through two experiments. First, a pilot study was conducted to determine the ecological validity of the intervention, identify potential practical issues associated with implementation in an elementary school setting, and gather feedback from teachers to make iterative improvements to the set of 4 games. Following a series of revisions to the intervention based on teacher feedback and examination of student learning from the pilot study, a preregistered experiment was conducted to estimate the causal effect of playing a set of 6 different games of Fraction Ball on student rational number learning.

Research Questions & Hypotheses

This study was guided by the following research questions:

1) Does playing Fraction Ball affect student math learning?

We hypothesized that students receiving the Fraction Ball intervention would have higher scores on all subtests and on the overall composite score of our measures of fraction and decimal arithmetic compared to the control condition, controlling for baseline scores.

2) Does student fraction and decimal number learning extend from the content presented in the games to near and far transfer rational number concepts?

We hypothesized that student performance on the aforementioned measures will be higher on near-transfer items than far-transfer items.

3) Is the impact of Fraction Ball on overall student learning moderated by grade, gender, or prior knowledge?

Although we did not make predictions that the impact of Fraction Ball would be moderated by other variables based on grade, gender, or prior knowledge, we explored possible interactions and treatment effects within each subgroup (e.g. treatment effect on subgroup of boys only) to probe the robustness of the treatment effect and for actionable insights into which groups (if any) were benefitting differentially from the intervention. This research question was motivated by feedback from teacher partners who expressed concern that this intervention might disproportionally benefit students who are older, boys (who they perceived to have on average more interest in basketball), and students who were already strong in math.

Pilot Experiment

Method

The Fraction Ball intervention was designed to use embodied cognition to increase children's rational number knowledge. The intervention consisted of a set of 4 games, each of which could be played during a 50-minute physical education (PE) class period. See Table S1 for details about the games and associated learning goals; the full script used by teachers is provided in the Online Supplementary Materials, Appendix A. The script details the rules of the game which are intentionally designed to reinforce rational number learning. Three participating 5th and 6th grade teachers co-designed the Fraction Ball court and script with the researchers to ensure that the math content aligned with their classroom instruction and their students' rational number knowledge. The games have different objectives like scoring as many points as possible in 2-minutes or racing to an exact number (e.g., 3.25) without overshooting the goal. Teams competed against each other, providing motivation for children to engage in fraction and decimal conversions and arithmetic, quickly and accurately, in a playful and physically active context.

Fraction Ball Court

As shown in Figure 1, the distance from below the basket to the 3-point arc is converted into a "0 to 1" area with arcs that act as number line markers and divide one end of the court into fourths, and the other end into thirds. Thus, the arcs closer to the basket represent 1/4, 1/2, and 3/4 point shots on one end of the court and 1/3 and 2/3 point shots on the opposite end. To promote fraction to decimal conversions, each end of the court is split in half with fraction and decimal symbols side by side (e.g. 1/4 0.25; 2/3 $0.\overline{66}$).[1] On the sideline of the court children tally

---

[1] We chose to represent 2/3 as $0.\overline{66}$ in response to teacher requests that the representation of this value match classroom instruction to prevent confusion and encourage students to say "0.66 repeating" as they do in the classroom.

their score on a 0-7 number line with fraction and decimal representations (Figure 1). During the games, children take on three roles —shooter, rebounder, and counter. The shooter takes a shot, the rebounder calls out the value of the shot, and the counter moves down the number line to tally the score.

Participants

All study procedures were approved by the BLINDED University IRB, protocol # BLINDED. Participants were 69 children (35 5th graders and 34 6th graders), and four teachers: one 5th grade, two 6th grade, and one STEM coordinator from a school in Santa Ana, California. The grade levels were decided by our teacher partners based on which grades they felt would benefit the most and align best with their curricular instruction. Children (45% female) were predominantly low-income Latino (over 90% in the school as indexed by qualification for free or reduced price lunch). In all participating classrooms, the students that provided assent within the same classroom were matched on gender and whether they scored above or below the average standardized pretest score, representing a total of 4 blocks (above average/below average and girls/boys). Half of the students in each block were randomly assigned to participate in the Fraction Ball intervention during PE while the other half remained in their usual PE. Descriptive statistics are provided for all measured child-level variables split by treatment group in Table S2. Two participants did not complete 90% or more of their untimed posttest items. They were both in the control group, so their inclusion would increase treatment effect estimates. We chose to drop them from our preferred specification (conservatively assuming nonresponse was not affected by the intervention), but report estimates including them in the online supplementary materials. One classroom did not fully comply with randomization. Instead, the teacher selected students, attempting to balance groups on the teacher's judgments of the students' math skills and gender. Pretest scores are similar between the treatment and control group whether we only include classrooms that adhered to random assignment or include this classroom. However, because assignment was not truly random in this classroom, we report estimates separately using the full sample (N = 69) and from the compliant sample (N = 54).

Procedure

Fraction Ball sessions took place twice a week for 2 weeks and were run by two teachers (one on each end of the court) within a 50-minute class period with 16 students per session. The duration of the Pilot Experiment was also collaboratively decided on with our teacher partners to strike a balance between what would be feasible given their demanding schedule and what we thought might be a reasonable amount of exposure to the intervention to see impacts. Because of time for student transitions and introductions to the games, sessions lasted between 35 to 40 minutes. All teachers attended a single 50-minute training session guided by the research team prior to implementation where they played and reviewed Fraction Ball games and were provided a script for games. Teachers were instructed to follow the script to ensure that students played the games in the same order and had a similar experience across groups. Fidelity was calculated via observational protocols. All children completed two math assessments within a week of the intervention starting and ending. Each testing session lasted approximately 20-30 minutes.

Measures

Fraction and Decimal Skills

Fraction and decimal knowledge were measured with a 50-item assessment with both timed (20-items) and untimed (30-items) sections. The measure was designed to assess a broad set of rational number skills related to children's rational number magnitude understanding, arithmetic, conversions between fractions and decimals and vice versa, and automaticity. The timed section aimed to assess efficient processing of rational numbers as a result of repeated play, whereas the untimed section aimed to capture knowledge of rational number concepts and procedures. The timed and untimed sections consisted of the following types of problems that comprised 8 mutually exclusive subtests (3-subtests within the timed section and 5-subtests within the untimed section). Subtests were designed to provide insight into the types of knowledge Fraction Ball was impacting, as well as identifying areas for improvement. A Near Transfer composite (24-items), measured proximal knowledge on procedures and numbers that were presented during the game. For example, identifying 1/4 or 0.25 on a number line would be classified as a near-transfer item. A Far Transfer composite (26-items), included numbers and/or skills that were not directly presented during the game. For example, identifying 5/10 or 0.80 on a number line would be classified as far transfer items, because students would not have had the opportunity to memorize these magnitudes from the Fraction Ball court; thus, higher performance on such items in the Fraction Ball condition than the control condition indicates impacts on some kind of generalizable knowledge. The Near and Far Transfer composites were composed of all of items from the subtests, and the two composites had no item overlap. Each item was scored as correct or incorrect and we calculated the average score for each participant. Fractions did not have to be reduced to their simplest form, and decimals had to represent the exact value (no rounding). No partial points were given, and unattempted items were marked incorrect. Identical versions were administered at pretest and posttest.

Timed Subtests. Thirty items were designed to capture students' ability to quickly solve fraction and decimal arithmetic problems and to convert between fractions and decimals within a 3-minute period. These items comprised the 3 timed subtests: a) fraction addition (10-items; same and different denominators, mixed fractions, and improper fractions), b) converting fractions to decimals (10-items), and c) converting decimals to fractions (10-items; an approach to measuring fraction arithmetic similar to Siegler et al., 2011).

Items from each subtest were presented in an interleaved sequence (see Online Supplementary Materials, Appendix B). All subtests were mean centered and standardized by dividing by the average of the standard deviations (SDs) across each grade-level and using only the control group SDs on posttest scores. Thus, impacts can be interpreted as grade standard deviations. Internal consistency was high for the pretest $\alpha = 0.93$ and the post-test $\alpha = 0.91$.

Untimed Subtests. Twenty items were designed to capture students' procedural and conceptual knowledge of rational numbers through 5 subtests presented in blocked sequence: a) fraction number line (FNL) estimation (5-items) on a 0-to-1 number line, b) FNL estimation (4-items) on a 0-to-5 number line (Siegler et al., 2011), c) fraction to decimal conversion (4-items), d) decimal to fraction conversion (4-items), and e) adding fractions to decimals (3-items; e.g. 2/4 + 0.25).

Scoring of Number Line Items

Accuracy for the FNL items was scored using percent absolute error (PAE; Booth & Siegler, 2006), where PAE = [absolute value of (student's response – correct response)]/(size of number line). If students left a FNL item blank and the student did not miss 90% or more of their untimed posttest, we replaced the missing value with the 90th percentile of PAE values given by students who answered the item. Because PAE scores are often positively skewed, the PAEs were transformed on a natural log scale and reverse coded so higher values represent greater accuracy.

Fidelity of Implementation

To capture the extent to which teachers administered the intervention in the way it was intended, we observed each of the intervention sessions using a fidelity rubric. Two researchers (who included, at various times, undergraduate and graduate student research assistants and the authors) marked the completion of each activity within the games (see Table S1 for examples) and recorded the duration of each game with a stopwatch. 13% of observations were double-coded and all coders established greater than 80% interrater reliability prior to coding independently. Teacher fidelity to each activity within the games ranged from 75%-100% (Table S6).

Results & Discussion

Descriptive statistics are presented in Table S2 and descriptive statistics of standardized composite scores are presented in Table S3. We conducted separate OLS regression analyses on standardized pretest scores (criterion) and treatment (predictor). We found no statistically significant differences between groups at pretest (see Figure S1 for group differences at baseline). To probe the robustness of our estimates across different assumptions, we report three sets of impact estimates: a) before (full sample no covariates, N = 68) and b) after statistically controlling for pretest scores (full sample with covariates, N = 66), and c) controlling for pretest scores and excluding the classroom of students that did not comply with random assignment and/or missed 90% or more of the untimed posttest (compliant sample with covariates, N = 54; see Table S4). Covariates were a composite z-score of child performance on the pretests and grade level. The pattern of impacts was consistent across specifications and is reported in Tables S4 & S8. Figure 2 shows the estimates from our preferred specification (pretest controls, complying classrooms only).

Impact estimates on the posttest composite score were consistently between .1 and .2 SD and not statistically significant ($.15 < ps < .30$). Across subtests, the only statistically significant impact was on converting fractions to decimals ($\beta = .55$, $p = .01$). Impacts on all other subtests were not significant ($-.18 < \beta < .27$; $.14 < ps < .99$; see Table S4).

As seen in Figure S2, there was a trend toward boys and students with higher pretest scores being positively impacted by playing Fraction Ball. However, as summarized in Table S8, we did not find any significant interaction effects between treatment and our three moderators.

Taken together, findings suggested that playing Fraction Ball improved children's skill at converting fractions to decimals, but had little impact on other kinds of fraction knowledge. Impacts were near zero on the overall composite score, on number line items, and on fraction addition items in both timed and untimed tests. Still, we were encouraged by our observations of how children played Fraction Ball, and thought the context provided several affordances that we could more efficiently capture to improve children's fraction knowledge. During our naturalistic observations of the games and semi-structured interviews, we learned that students and teachers alike, showed enthusiasm about the game while they were playing and after. Teachers reported that students were asking them when they could play Fraction Ball again. We also received anecdotal reports from the teachers that students who were in the Fraction Ball condition became leaders in the classroom during fraction instruction activities explaining concepts to their classmates who were not assigned to the intervention. Thus, although we did not view the Pilot Experiment as definitive evidence for the usefulness of Fraction Ball, we believed that it showed enough promise to justify efforts toward improving it further. To strengthen the intervention, we engaged in one focus group session and 11 interviews with teachers which provided fruitful insights for improvements. For example, teachers noticed that many children were simply walking silently down the number line to keep track of their score, which does not give the opportunity for teachers to hear students' thinking and may enable the use of strategies that are less likely to transfer to traditional mathematical settings. Teachers also developed ideas to connect the games to classroom instruction by using "number talks" (a think aloud strategy used in the school) and recording students' own explanations of their mathematical reasoning. We also examined the posttest scores by subcategories and identified number line estimation and far transfer items (i.e., fraction representations that were not displayed on the court) as areas for improvement. Iterations based on these observations are described in the next section below. We hypothesized that with additional experience and development of the games, as well as an increase in the amount of time children spent engaging with the court, Fraction Ball had potential to generate larger and broader impacts on children's rational number skills. We tested this hypothesis in the Efficacy Experiment.

Efficacy Experiment

The Efficacy Experiment was designed to incorporate feedback from the teachers and research team from the Pilot Experiment to improve Fraction Ball, and to evaluate the effectiveness of Fraction Ball in a sample with adequate power (using a better-informed benchmark) and increased dosage. For the Efficacy Experiment, we made three categories of modification. First, teachers required students on the number line to state the problem they were completing out loud (e.g., "one fourth plus three fourths equals one whole") so students would be more likely to connect their actions to the corresponding mathematical concepts and so teachers could identify students who were struggling and provide support. Second, we added debriefing activities on the number line for teachers to recap what happened after each round of the games and reinforce concepts that they noticed students grappling with.

Third, we added additional components to the games to target number line estimation and different fraction representations (e.g., improper fractions or denominators not represented on the court). To target number line estimation, we added a “Ghost Number Line” drawn in chalk with a 0 at the beginning and 5 at the end with no notation or markings in between. When playing on the Ghost Number Line students had to estimate the value of each shot scored without the scaffolding of the markings on the number line. Additionally, scoring the most points was not the way to win the game, instead it was the team who most accurately estimated their total score on the Ghost Number Line that won. The number line was 10-feet long so each 2-foot segment was 1 unit meaning that students could use a measuring tape to measure exactly where they were supposed to be on the number line after they estimated and see which team was closer. We expected the Ghost Number Line would improve number line estimation as it would force students to mentally and repeatedly estimate their score without the hash-mark scaffolds of the number line.

The next additions were “Supercharge Points” and “Who goes first?” which allowed teachers to introduce different fraction representations into the game (e.g., improper fractions like 6/4, or different denominators like 5/10). At any point in the game the teacher could yell Supercharge Points and the game would freeze, the teacher would say a number (from a list of numbers we co-created with the teachers) and the students on the number line would have a chance to add that number to their score. If they added correctly, they were allowed to keep the points, if they added incorrectly, they had to return to where they were on then number line. For “who goes first” teachers would call a number (from the same list which included improper fractions and unique denominators) before the game began and the counters would race to that number on the number line, the team that got there first shot first in the game. We hoped this would help students receive more practice with improper fractions and mixed numbers that were not represented on the court.

Research Questions & Hypotheses

The research questions, hypotheses, and analyses for the Efficacy Experiment were the same as for the Pilot Experiment and were pre-registered at https://osf.io/6qysh after the pretests were administered but prior to data entry and analysis. Not all preregistered hypotheses are addressed within this current manuscript, because a thorough analysis of the language interactions is in preparation and will be reported elsewhere. All preregistered analyses related to student learning are reported in the current article. Together, these two articles will report all preregistered analyses.

Method

Participants

Participants were 10 teachers and 16 students from each of their classrooms: four 4th-grade, four 5th grade, and two 6th grade classrooms. As in the Pilot Experiment, students that provided assent within the same classroom (n = 195 of 232; 16-25 per class) were matched on gender, to create a total of 2 blocks per classroom, 1 block for boys and 1 block for girls. Eight students in each gender block were randomly assigned to the Fraction Ball intervention during PE while the other half remained in their usual PE, this randomization yielded a total of 8 children for each treatment condition (8 x 2=16) within each class (16 x 10) making our full analytic sample 160 students. The remaining 35 students that were not randomized where excluded from the analytic sample to ensure equal group sizes and minimize the cost of data collection. Our full sample consisted of 160 students; however, seven students were excluded for: being absent on posttest (2 treatment), being absent during pretest (3 treatment) or failing to complete 90% or more items of the untimed posttest (2 control). We estimate effects with and without these participants (see Table S5). Due to involvement in other school activities three students did not comply with being assigned to receive the Fraction Ball intervention, however, these students completed pretest and posttest, thus we included students in the analysis sample to estimate the effect of being assigned to the treatment. There were 7 students who participated in the Pilot Experiment in 5th grade that also participated in the Efficacy Experiment when they were in 6th grade. In the Pilot Experiment, 3 students were in the control condition and 4 students received the treatment. In the Efficacy Experiment, 5 students were in the control condition and 2 students received the treatment. The study was administered in entirety to half of the classrooms first, and then to the second half of classrooms to accommodate scheduling on one Fraction Ball court. Descriptive statistics are provided for all measured child-level variables split by treatment group in Table 1.

INSERT TABLE 1 HERE

Procedure

Pretests were administered to all the students one week prior to starting the intervention by their teachers. The students in the intervention played two games per week for 50 minutes across three weeks, for a total of six games in 300 minutes. Six classrooms completed their sessions during the fall semester and the remaining four classrooms completed 3 sessions during the fall semester and 3 sessions after winter break. Despite this delay all classrooms completed the same six sessions. The posttest measures were administered to students within one week following the end of the intervention.

Measures & Fidelity

The Efficacy Experiment used the same measures as the Pilot Experiment. Measures had high internal consistency: for the Timed Subtest (α = 0.91 and α = 0.93) and the Untimed Subtest (α = 0.88 and α = 0.91) for the pretests and post-tests respectively. Previously trained observers marked the completion of each activity within the games and recorded the duration of each game with a stopwatch. Teacher fidelity to each activity within the games ranged from 52%-96% (Table S7).

Results & Discussion

In our pilot study, we included a pretest and recruited as many participants in the target grades as we could. The minimum detectable effect size calculated for a two-sided hypothesis test at the .05 significance level at 80% power is 2.8 times the standard error (Bloom, 1995). Thus, in our preferred specifications in the Pilot Experiment, we were powered to detect an impact on the posttest composite of 2.8 times .11 or .12 (the standard errors of our estimated treatment impacts in those specifications), which is approximately .31. Since the estimated impacts from the Pilot Experiment in those specifications were .12 and .18, we did not have sufficient power post-hoc to detect a treatment impact on the posttest composite score for the Pilot Experiment. However, post-hoc power is a volatile measure, and we had other information available: First, our priors were that the improved intervention would be slightly more effective than this, even in the Pilot Experiment. More importantly, we forecasted that impacts in the efficacy study would be larger than impacts in the pilot study. Thus, given our larger sample size, we viewed the efficacy study to be adequately powered to detect the primary preregistered impact. Based on the standard error from our preferred specification of our efficacy study, the study was powered to detect an impact of 2.8 times .11, which is .31 SD, when the observed impact from our preferred specification was .44 SD.

To probe the robustness of our estimates across different assumptions, we report three sets of impact estimates: a) before (full sample no covariates, N = 158) and b) after statistically controlling for student grade and pretest scores (full sample with covariates, N = 155), and c) controlling for student grade and pretest scores and excluding the students that were missing 90% or more of the untimed posttest (full sample with covariates not missing 90% of post untimed test, N = 157, see Table S5). The pattern of impacts was consistent across specifications and is reported in Tables S5 & S8. Figure 2 shows the estimates from our preferred specification (pretest and grade controls, not missing 90% of post untimed test). Descriptive statistics are presented in Table 1 and standardized composite scores in Table 2. Similar to the Pilot Experiment, there were no significant differences between groups at pretest (Figure S1). The results of the regression analyses are summarized in Tables S5 & S8. Again, estimates were similar across levels of statistical control and exclusion criteria, suggesting that baseline imbalance was not a problem in this study.

INSERT TABLE 2 HERE

We found that students who played Fraction Ball outperformed the control group by .40 SD on their overall posttest composite score (β =.40), a moderate to large effect. Similar and consistent effects were seen for untimed conversions from decimals to fractions (β =.51, $p <$ .001), and both the timed and untimed conversions from fractions to decimals (β_{timed} = .48, $\beta_{untimed}$ =.69; all *p*s <.001). Impacts on number line 0 to 1 and 0 to 5 subtests were in the positive direction but not significant ($\beta_{0\ to\ 1}$ = .13, p = *.33;* $\beta_{0\ to\ 5}$ =.14, p = .35). Impacts were nonsignificant and positive for untimed fraction addition (β = .24, p = .15), but negative for timed fraction and decimal addition (β = -.13, p = .34). The impact of playing Fraction Ball was also moderate to large on students' Near Transfer knowledge (β =.42, p < .001). However, we did not find significant impacts on children's Far Transfer skills (β =.12, p = .37)

Similar to the Pilot Experiment, we did not find moderation effects between treatment and the three moderators (-.08 < βs < -.02, ps > .711) as reflected in Table S8. The effects of playing Fraction Ball were mostly consistent across subgroups with moderate to large significant impacts ranging between .4 and .6 SDs (Figure S2); treatment effect was directionally, although non-significantly smaller for fourth (β =.28) and 6th graders (β = .22).

In summary, Fraction Ball improved children's overall rational number ability (composite score), arithmetic that was most closely tied to the content on the court (near transfer items) and converting between decimals and fractions in timed and untimed settings. Revisions from the Pilot Experiment seem to have been successful in expanding effects to additional rational number skills.

General Discussion

This study demonstrated the ecological validity and learning value of a play-based, rational number learning intervention in a school that serves predominantly low-income Latino students. The intervention leveraged key principles of math and embodied cognition research to boost students' rational number learning using a playful and engaging approach. Fraction Ball produced moderate to large effects on students' rational number abilities using a gold standard randomized to groups pretest posttest experimental design. Specifically, Fraction Ball had a significant impact on the overall and near transfer composite scores, following iterative improvements made to the intervention between the Pilot and Efficacy Experiments. These findings demonstrate the power of playful learning shaped by research in cognitive development (Schlesinger et al., 2020).

The design and evaluation process of this study had several strengths which may be meaningful to future intervention researchers. First, the design and analytic plans were nearly identical across both experiments; we used an internal preregistration document in the Pilot Experiment and a preregistration on OSF in the Efficacy Experiment, limiting our researcher degrees of freedom. Second, the iterative improvement approach which included a feasibility study that informed the second efficacy experiment allowed us to improve the intervention and examine the effectiveness of Fraction Ball after new additions and iterations. In general, we noticed stronger impacts of Fraction Ball on students' overall knowledge and Near Transfer concepts in the Efficacy Experiment, which may have been a result of higher dosage, requiring students to verbalize their arithmetic calculations, and new games requiring students to engage with a wider range of rational number concepts.

Playing Fraction Ball had the greatest positive impacts on students' skills converting between fractions and decimals. A learning mechanism that could explain a significant portion of these impacts may be related to learning from comparisons because fractions and decimals were presented side-by-side on the Fraction Ball court in two key areas: a line that divided the court in half, and on the number line "scoreboard" with both areas showing equivalent fraction and decimal symbols on each side (see Figure 1). Also, the game dynamics and rules (for example see explanation for *Exactly-Flip* in Table S1) required students to switch between fraction and decimal notations mid-game and keep track of scores in both notations. These opportunities for comparing and contrasting fraction and decimal notations may have been key for students to notice similarities (e.g. equivalency) and differences (e.g. notation style) between fractions and decimals in a manner that promoted a more robust understanding of fraction and decimal equivalence. While this explanation is supported by theoretical accounts and evidence from the literature on learning from comparisons (Alfieri et al., 2013; Richland et al., 2017; Rittle-Johnson et al., 2017), our data and study design – as typical in field experiments – did not allow us to fully dismantle the intervention to isolate effective and ineffective components. Still, the stronger impacts in the Efficacy Experiment than in the Pilot suggest that teacher and experimenter feedback was useful for identifying levers for improving the intervention to achieve an effect size on our preregistered composite of .44 SD. It is difficult to compare effect sizes for interventions that use different researcher-designed measures (e.g., Kraft, 2020) and student populations. In future work, we hope to test the effectiveness of the fraction intervention by measuring impacts on broad standardized achievement measures.

However, we attempt to briefly contextualize the magnitude of this impact compared with prior relevant interventions. Fazio and colleagues' (2016) fraction number line intervention was less intensive than our study, including 15 minutes of training. Children in the treatment group improved by between .13 SD and .58 SD more from pretest to posttest relative to an active control group on a set of three fraction number tasks. This range of impacts is similar to those we report, although we would characterize the tasks used in the current study as somewhat more distal to the intervention (in format, setting, and timing). Still, there is significant room for improvement: In a 5-year set of experiments on a 12-week intensive fraction tutoring program targeted to students at risk for low math achievement, Fuchs and colleagues (2016) reported impacts ranging from approximately 1 SD on number line estimation, between 1 and 2.5 SD on fraction calculation, and between .4 and .9 SD on a set of relevant problems selected from the NAEP. The large number line estimation and fraction arithmetic impacts likely partially reflect strong overlap with the program content; however, these are important skills worth teaching directly. We hope that with increased dosage and integration into a strong rational numbers curriculum, Fraction Ball might improve student learning to a greater extent.

Notably, we did not find effects on subtests of far transfer items (i.e., arithmetic and estimation that was not represented on the court), number line estimation, and fraction/decimal addition in either experiment. In the Pilot Experiment effects were near 0 on far transfer items, number line estimation, and fraction/decimal addition. However, in the Efficacy Experiment we added new rules (e.g., stating math out loud and debriefing sessions after each round) and games (e.g., Ghost Number Line and Supercharge Points) to target these skills. While the treatment effects were not significant, effect sizes in the Efficacy Experiment were in the .2 to .3 range suggesting movement in the positive direction. One possibility is that far transfer items, number line estimation items, and arithmetic items, involved improper and mixed fractions and there is preliminary evidence that area model representations may be more helpful for learning than number line training in the short-run (Tian et al., 2021). Future well-powered work will be needed to test whether improved iterations of Fraction Ball can produce robust impacts on such items.

With respect to far transfer and number line estimation, it appears that eliciting transfer to more advanced rational number arithmetic and estimation may require more explicit instructional supports, higher dosage, and/or further refinements in game design by integrating principles from the science of learning and teacher/student feedback. For example, we are currently co-designing classroom activities with teachers, that reinforce and build on the concepts presented on the court. One activity involves watching a WNBA or NBA game during class and pausing the game after baskets are scored for students to estimate the value of that shot on a Fraction Ball court with different denominators (halves, thirds, quarters, sixths, etc.). This activity of repeated estimation can then be extended to ask students to tally all points made and engage in fraction addition of the estimated points of each team to understand which team won. In our ongoing work we are investigating many game improvements and avenues for Fraction Ball to be leveraged in classroom instruction. Additionally, research is needed to examine the durability of results through a delayed posttest, and explore if the impacts of Fraction Ball generalize to the kinds of broad academic achievement tests that drive policy discussions and decision making. Through continued iterative improvement and broadening the learning goals of the games to include more advanced content (e.g., rational number multiplication and division, or probabilities), we hope to answer some of these questions through larger scale evaluations.

This study serves as a model for how cognitive psychological research can inform intervention design, how iterative intervention implementation ensures usability, and how evidence-based and ecologically valid interventions can improve student learning. Our study used an experimental design where we randomized students within classrooms to either be in the intervention or business as usual control group and included a range of robustness checks to rule out threats to internal validity, making it strong with respect to internal validity. However, there are limitations to the external validity of our study as results may not generalize to other student populations (e.g., different age groups or regions of the country or world), settings (e.g., after school programs or public parks), and intervention approaches (e.g., including additional math content or not having teachers present to facilitate). Most importantly, children in both experiments participated in regular PE class in the control group, and both groups received their regular mathematics instruction. Therefore, the current studies should not be used as evidence of the effectiveness of Fraction Ball as a substitute for regular mathematics instruction, but rather should be viewed as a supplement.

Future studies should try to understand the relative effectiveness of each game and their components to better understand the mechanisms through which the intervention works and to provide insights into how to improve it further. As outlined in the introduction, Fraction Ball game design and rules draw from multiple literatures with implications for understanding the contribution of each. For example, the contribution of using comparisons could be tested by comparing the current design to a court which uses only fractions on one half of the court and only decimals on the other half of the court, instead of both representations on the same side of the court. Also, the contribution of the linearity of the court and scoreboard designs based on the Integrated Theory of Numerical Development could be tested by comparing the current design with a court which uses a non-linear point system (e.g., further away from basket could be more or less points). Data from Ramani & Siegler (2008, 2011), suggest this could be an essential visuospatial cue since in their studies, children who played with a circular version of their game did not show benefits. The contribution of whole body movement could also be examined by comparing outdoor play of Fraction Ball with a paper & pencil or digital app version of Fraction Ball that limits whole body movement. As previously noted by Ramani & Siegler (2009), positive effects from game play which incorporates multiple cues, e.g., visuospatial, kinesthetic, auditory, and temporal cues may be related to the redundancy of cues where not a single cue is essential, but there may be an effect similar to parallel distributed processing models of speech perception, reading, and other skills where there this a collective effect as cues reinforce one another (McClelland & Elman, 1986). As mentioned above, in Ramani & Siegler (2008, 2011), linearity could be essential for students to benefit from this collective effect since it could serve as a visual representation that integrates embodied cues to form a conceptual schema which reinforces magnitude understanding of fractions/decimals.

We encourage future laboratory and field experimentation from researchers and educators outside of our research group to test or implement Fraction Ball. As noted, the intervention materials used in the Efficacy Experiment are freely available, and a Fraction Ball court can be drawn on an existing outdoor basketball court with chalk or paint. Future studies should engage students to incorporate their feedback for game improvements and to gauge the extent to which students found the game fun, motivating, valuable, and whether it has an impact on their attitudes towards rational numbers. Another important avenue that could lead to greater improvements for students would be to ensure that teachers aligned their curriculum and time of rational number instruction with the Fraction Ball intervention. Future work should also explore expanding Fraction Ball to public park settings with light signage or QR codes to share background and rules for games and observe how families use the court in a more naturalistic context. Critically, children and teachers were enthusiastic about Fraction Ball and engaged in rich math-oriented discussion and interactions while having fun and exercising. Fraction Ball represents a low-cost, highly scalable intervention (the materials from the version tested in the Efficacy Experiment are freely available on OSF) that could promote rational number learning through a playful and engaging approach in schoolyards and public parks around the world.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*(2), 87-113. https://doi.org/10.1080/00461520.2013.775712

Barbieri, C. A., Rodrigues, J., Dyson, N., & Jordan, N. C. (2020). Improving fraction understanding in sixth graders with mathematics difficulties: Effects of a number line approach combined with cognitive learning strategies. *Journal of Educational Psychology*, 112(3), 628.

Bialystok, E., Craik, F. I., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663.

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation review*, 19(5), 547-556.

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology*, *37*(4), 247–253.

Bottge, B. A., Ma, X., Gassaway, L., Toland, M. D., Butler, M., & Cho, S. J. (2014). Effects of blended instructional models on math performance. *Exceptional Children*, 80(4), 423-437.

Braithwaite, D. W., & Siegler, R. S. (2020).Putting fractions together. *Journal of Educational Psychology*. Advance online publication. https://doi.org/10.1037/edu0000477.

Bustamante, A. S., Schlesinger, M., Begolli, K. N., Golinkoff, R. M., Shahidi, N., Zonji, S., Riesen, C., Evans, N., & Hirsh-Pasek, K. (2020). More than just a game: Transforming social interaction and STEM play with Parkopolis. *Developmental Psychology, 56*(6), 1041–1056. https://doi.org/10.1037/dev0000923

Common Core State Standards Initiative. (2010). *Common Core State Standards for Mathematics*. http://www.corestandards.org/

Dyson, N. I., Jordan, N. C., Rodrigues, J., Barbieri, C., & Rinne, L. (2020). A fraction sense intervention for sixth graders with or at risk for mathematics difficulties. *Remedial and Special Education*, 41(4), 244-254.

Fazio, L. K., Kennedy, C. A., & Siegler, R. S. (2016). Improving children's knowledge of fraction magnitudes. *PloS one*, *11*(10), e0165243. https://doi.org/10.1371/journal.pone.0165243

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Malone, A. S., Wang, A., Hamlett, C. L., Jordan, N. C., Siegler, R. S., & Changas, P. (2016). Effects of intervention to improve at-risk fourth graders' understanding, calculations, and word problems with fractions. *The Elementary School Journal*, 116(4), 625–651.

Fuchs, L. S., Malone, A. S., Schumacher, R. F., Namkung, J., & Wang, A. (2017). Fraction intervention for students with mathematics difficulties: Lessons learned from five randomized controlled trials. *Journal of Learning Disabilities*, 50(6), 631-639.

Givvin, K. B., Stigler, J. W., & Thompson, B. J. (2011). What community college developmental mathematics students understand about mathematics, Part II: The interviews. *MathAMATYC Educator, 2*(3), 4–18.

Hamdan, N., & Gunderson, E. A. (2017). The number line is a critical spatial-numerical representation: Evidence from a fraction intervention. *Developmental Psychology*, 53(3), 587.

Hansen, N., Jordan, N. C., & Rodrigues, J. (2017). Identifying learning difficulties with fractions: A longitudinal study of student growth from third through sixth grade. *Contemporary Educational Psychology*, *50,* 45-59. https://doi.org/10.1016/j.cedpsych.2015.11.002

Hassinger-Das, B., Bustamante, A., Hirsh-Pasek, K., & Golinkoff, R. M. (2018). Learning Landscapes: Playing the Way to Learning and Engagement in Public Spaces. *For Peer Review*, *8*, 1–46. https://doi.org/10.3390/educsci8020074

Hassinger-Das, B., Toub, T. S., Zosh, J. M., Michnick, J., Golinkoff, R., & Hirsh-Pasek, K. (2017). More than just fun: A place for games in playful learning/Más que diversión: El lugar de los juegos reglados en el aprendizaje lúdico. *Infancia y Aprendizaje*, *40*(2), 191-218.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest*, 16(1), 3-34. https://doi.org/10.1177/1529100615569721

Link, T., Moeller, K., Huber, S., Fischer, U., & Nuerk, H. (2013). Trends in Neuroscience and Education Walk the number line – An embodied training of numerical concepts. *Trends in Neuroscience and Education*, *2*(2), 74–84. https://doi.org/10.1016/j.tine.2013.06.005

Malone, A. S., Fuchs, L. S., Sterba, S. K., Fuchs, D., & Foreman-Murray, L. (2019). Does an integrated focus on fractions and decimals improve at-risk students' rational number magnitude performance? *Contemporary Educational Psychology*, 59, 101782.McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology*, 18(1), 1-86.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics* (Vol. 1). Author.

NGSS Lead States. (2013). *Next Generation Science Standards: For States, by States.*

Richland, L. E., Begolli, K. N., Simms, N., Frausel, R. R., & Lyons, E. A. (2017). Supporting mathematical discussions: the roles of comparison and cognitive load. *Educational Psychology Review*, *29*(1), 41-53. https://doi.org/10.1007/s10648-016-9382-2

Richland, L. E., Stigler, J. W., & Holyoak, K. J. (2012). Teaching the conceptual structure of mathematics. *Educational Psychologist*, *47*(3), 189-203. https://doi.org/10.1080/00461520.2012.667065

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*(2), 346.

Rittle-Johnson, B., Star, J. R., & Durkin, K. (2017). The power of comparison in mathematics instruction: Experimental evidence from classrooms. In *Acquisition of complex arithmetic skills and higher-order mathematics concepts* (pp. 273-295). Academic Press

Sarihi, N., Pournesaei, G. S., & Nikakhlagh, M. (2015). Effectiveness of group play therapy on behavior problems in preschool children. *Journal of Analytical-Cognitive Psychology*, *6*(23), 35-41.

Schlesinger, M., Hassinger-Das, B., Zosh, J. M., Sawyer, J., Evans, N., & Hirsh-Pasek, K. (2020). Cognitive Behavioral Science behind the Value of Play: Leveraging Everyday

Experiences to Promote Play, Learning, and Positive Interactions. *Journal of Infant, Child, and Adolescent Psychotherapy*, *19*(2), 202–216.

Schumacher, R. F., Jayanthi, M., Gersten, R., Dimino, J., Spallone, S., & Haymond, K. S. (2018). Using the number line to promote understanding of fractions for struggling 5th graders: A formative pilot study. *Learning Disabilities Research & Practice*, 33(4), 192-206.

Seethaler, P. M., Fuchs, L. S., Star, J. R., & Bryant, J. (2011). The cognitive predictors of computational skill with whole versus rational numbers: An exploratory study. *Learning and Individual Differences*, *21*(5), 536-542. https://doi.org/10.1016/j.lindif.2011.05.002

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early Predictors of High School Mathematics Achievement. *Psychological Science*, *23*(7), 691–697. https://doi.org/10.1177/0956797612440101

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, *11*(5), 655-661. https://doi.org/10.1111/j.1467-7687.2008.00714.x

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive psychology*, *62*(4), 273-296.

Stigler, J. W., Givvin, K. B., & Thompson, B. J. (2010). What community college developmental mathematics students understand about mathematics. *MathAMATYC Educator*, 1(3), 4-16.

Tian, J., Bartek, V., Rahman, M. Z., & Gunderson, E. A. (2021). Learning Improper Fractions with the Number Line and the Area Model. *Journal of Cognition and Development*, 22(2), 305-327.

Vamvakoussi, X., & Vosniadou, S. (2010). How many decimals are there between two fractions? Aspects of secondary school students' understanding of rational numbers and their notation. *Cognition and Instruction*, *28*(2), 181-209. https://doi.org/10.1080/07370001003676603

Zosh, J. M., Hirsh-Pasek, K., Hopkins, E. J., Jensen, H., Liu, C., Neale, D., & Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology*, *9*, 1124.

Table 1

Summary Statistics of Fraction Ball Moderators and Raw Scores of Outcomes for the Efficacy Experiment

	Full Sample			Control			Treatment			
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	*p*
Compliance	160	98%		83	96%		77	100%		.09
Male	160	49%		83	46%		77	52%		.44
4th grade	160	41%		83	41%		77	42%		.94
5th grade	160	40%		83	40%		77	40%		.95
6th grade	160	19%		83	19%		77	18%		.86
Dosage (in days)							77	5.92		
Participated in Pilot Experiment				5			2			
Attrition from Pre-test to Post-test	160	1%		83	0%		77	3%		.14
Missing Pre-test	160	2%		83	0%		77	4%		.07
Missing Posttest	160	1%		83	0%		77	3%		.14
Missing 90% not-timed post-test items	160	1%		83	2%		77	0%		.17
Pretests										
Timed fraction to decimal conversion	157	9%	15%	83	8%	13%	74	10%	16%	.39
Timed decimal to fraction conversion	157	18%	25%	83	18%	26%	74	18%	25%	.95
Timed fraction addition	157	46%	34%	83	47%	35%	74	44%	33%	.61
Untimed Fraction to Decimal Conversion	157	20%	28%	83	20%	28%	74	19%	28%	.78
Untimed Decimal to Fraction Conversion	157	37%	41%	83	41%	42%	74	32%	38%	.17
Untimed Fraction and Decimal Addition	157	24%	28%	83	24%	28%	74	23%	28%	.66
PAE 0 to 1	157	11%	12%	83	11%	11%	74	12%	13%	.64
PAE 0 to 5	157	32%	18%	83	33%	18%	74	31%	18%	.55
Post-tests										
Timed fraction to decimal conversion	158	19%	24%	83	14%	21%	75	25%	26%	.003**
Timed decimal to fraction conversion	158	28%	29%	83	25%	28%	75	30%	29%	.29
Timed fraction addition	158	50%	34%	83	52%	35%	75	48%	33%	.46
Untimed Fraction to Decimal Conversion	158	38%	39%	83	28%	33%	75	49%	43%	<.001***
Untimed Decimal to Fraction Conversion	158	48%	42%	83	41%	41%	75	56%	42%	.03*
Untimed Fraction and Decimal Addition	158	28%	31%	83	25%	30%	75	31%	31%	.27
PAE 0 to 1	158	11%	11%	83	11%	12%	75	10%	11%	.51
PAE 0 to 5	158	27%	18%	83	29%	18%	75	25%	18%	.23

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. p-value is based on a two-tailed t-test comparing treatment and control groups on each outcome. Raw scores were standardized using average grade standard deviations in the control group. Indices are only shown as standardized scores to facilitate interpretation as indices contain raw scores and natural log transformed percent average error scores from the number line items. PAE = percent absolute error. The N in the Variable section refers to total sample possible, including student attrition.

Table 2

Summary Statistics of Fraction Ball Standardized Scores of Outcomes for the Efficacy Experiment

	Full Sample			Control			Treatment			
Construct	*N*	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*p*
Pretests										
Average Score	157	-0.03	1.20	83	-0.00	1.18	74	-0.06	1.22	.78
Near Transfer	157	0.00	1.15	83	0.02	1.14	74	-0.02	1.17	.86
Far Transfer	157	0.00	1.27	83	-0.03	1.20	74	0.03	1.35	.78
Post-Tests										
Average Score	158	-0.00	1.21	83	-0.18	1.14	75	0.20	1.27	0.05*
Near Transfer	158	0.00	1.18	83	-0.20	1.09	77	0.22	1.25	0.03*
Far Transfer	158	0.00	1.25	83	-0.09	1.19	77	0.09	1.30	.37

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. p-value is for based on a two-tailed t-test comparing of each outcome by treatment group. Teacher ratings were collected from teachers based on their own criteria of math ability for each student in comparison to other students in the same class. Raw scores were standardized using average grade standard deviations in the control group. Indices are only shown as standardized scores to facilitate interpretation as indices contain raw scores and natural log transformed percent average error scores from the number line items. PAE = percent absolute error.

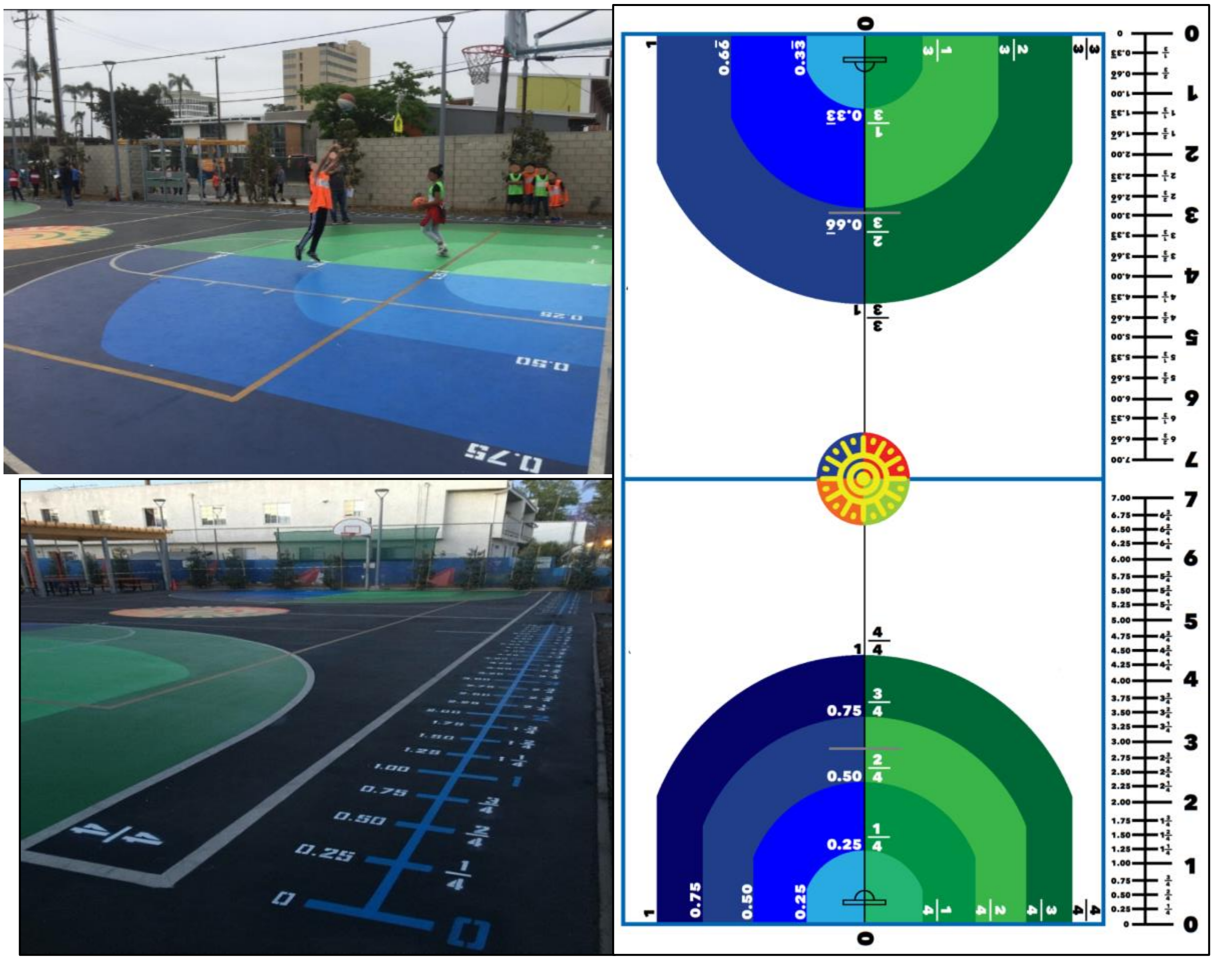


Figure 1. Fraction Ball court design.
Top Left: Structured games where children perform shooter, rebounder, and counter roles.

Bottom Left: Number line to assist children keeping track of their score.
Right: Full court graphic with a number line ranging from 0 to 1 and arches with fractions on right side of the court and a mirror image with decimals on the left side of the court.

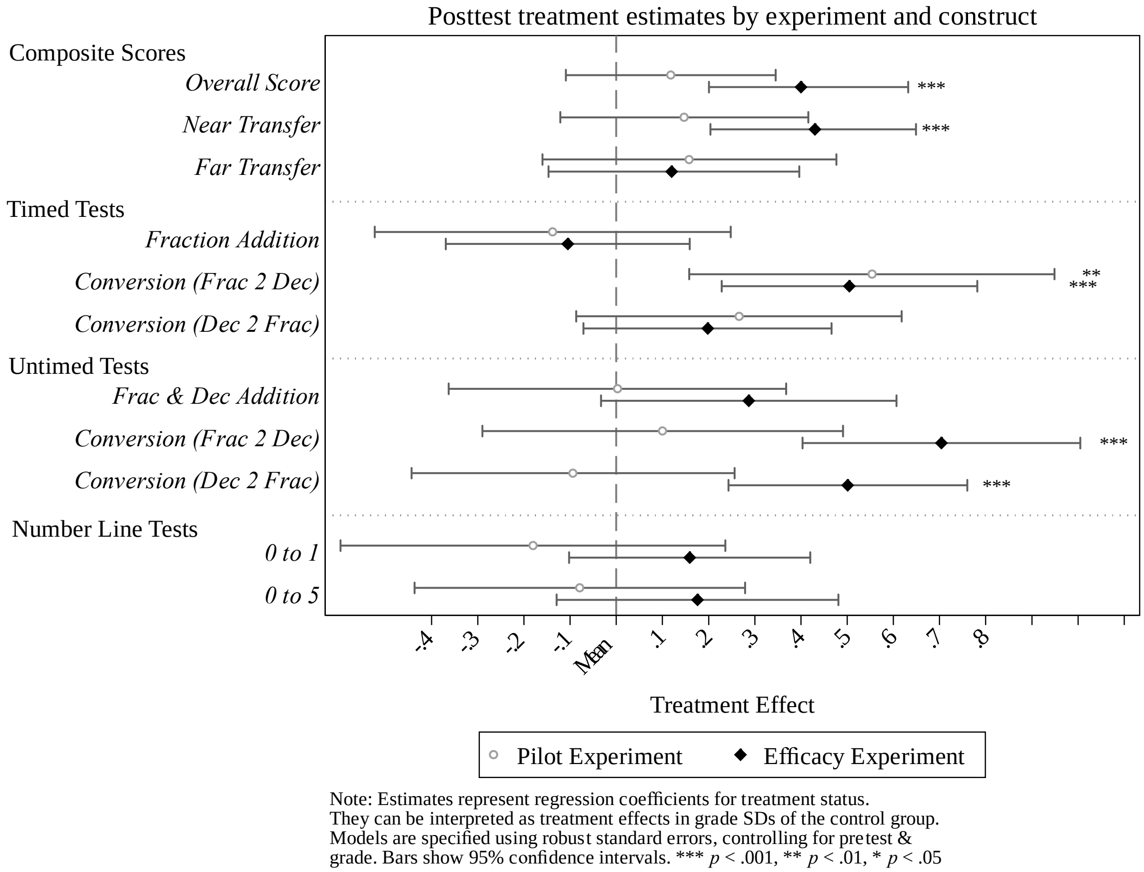


Figure 2. The effects of playing Fraction Ball for each construct and subtest for the Pilot and Efficacy Experiments.

Supplementary Materials

Standardizing Assessment Subtests

We standardized each subtest to create the 5 subtests, a)-e), in the same manner as the timed test, above. Internal consistency was high for the pretest α = 0.90 and the post-test α = 0.93, for Overall Knowledge, Near and Far Transfer Composite Scores. The Overall Score was a standardized composite of 8 standardized subtests from the timed and untimed test (50-items total). To create the Near Transfer and Far Transfer composite score we first created 8 parallel standardized subtests (as above), but only drew items from each subtest that contained fractions and decimals that were also present in the game of Fraction Ball for the Near Transfer construct (24-items) or not present in Fraction Ball for the Far Transfer construct (26-items) without overlap between the two measures. All three composite measures were mean centered and standardized by dividing by the average of the SDs for each grade and using control group SDs only for posttest scores. Therefore, the composite scores can be interpreted as grade-SD units.

Table S1

A summary of court design and game rules as well as their association to rational number skills.

Court Design	**Rational number target skills**
Fraction/Decimal scoring system.	(C) Conceptual fraction/decimal arithmetic Players have to add/subtract their score using fraction/decimal representations, thus, promoting arithmetic skills. Further, without any aids to follow algorithms, players are naturally prompted to develop a conceptual understanding of fraction/decimal arithmetic.
Points are based on magnitudes along the number line. Further is a larger fraction/decimal point.	(A) fraction/decimal magnitude representation (B) reduce "whole number" bias As players move further or closer the fraction/decimal representations increase/decrease, thus, providing an embodied experience of magnitude representations of each fraction/decimal and reducing whole number biases (e.g. 1 is smaller than .75 because it looks like a smaller number)
The center of the court is split to show fraction and decimal equivalency. Also, the number line used as a scoreboard has decimals on one side and fractions on the other side.	(D) conversion between fractions and decimals and (E) fraction/decimal comparison Players are able to see that 3/4 is equal to .75 as they are represented next to each other on the court and the number line.
General Game Rules	
Rapid Fire Players take turns to shoot and score as many points as possible in 1 minute.	(F) Automaticity Students must work quickly, increasingly relying on automatic fact retrieval in order to add fraction/decimal magnitudes in a running total as each shot is made.
Make it Count Players take turns and have a limited number of shots (3-shots per player) to make the maximum number of points	(C) Conceptual fraction/decimal arithmetic, (D) conversion between fractions and decimals, and (E) fraction/decimal comparison As teams get ahead of each other, the players on the losing team have to think strategically about the fraction/decimal magnitude from which to shoot, in order to catch up with the winning team. The team score comparison requires fraction to decimal conversion to know who’s ahead and by how much.
Exactly Players take turns to get an exact score in fraction/decimal magnitude (e.g. 5/4 or 1.25). The first team to reach the exact score wins. If both teams "overscore" then the team which is closest to the target wins.	(A) fraction/decimal magnitude representation and (C) Conceptual fraction/decimal arithmetic Similarly to *Make it Count*, players have to consider which fraction/decimal will bring them closer to the target score without "overscoring," thus, requiring teams to strategically consider the difference between the fraction/decimal magnitude of their current score and the magnitude of the target score in order to decide where to shoot from.
Note: The following games were added to the Efficacy Experiment after the Pilot	

Rapid Fire Target In "Rapid Fire Target" we combined the rules from "Rapid Fire" and "Exactly," such that each player has 1-minute to shoot and bring their team's score closer to a target fraction/decimal magnitude.	(A) fraction/decimal magnitude representation, (C) Conceptual fraction/decimal arithmetic, and (F) Automaticity See *Rapid Fire* & *Exactly* above
Make it Count–Ghost In the Ghost version of *Make it Count,* we added a "ghost" 0 to 7 number line (scoreboard) on the side of the court which does not have the fraction/decimal hatch marks. The team which has the closest alignment between their score and their estimation wins.	(A) fraction/decimal magnitude representation, (C) Conceptual fraction/decimal arithmetic, (D) conversion between fractions and decimals, and (E) fraction/decimal comparison Players have to engage in estimation of fraction/decimal representations as well as arithmetic using only the 0 and 7 endpoints on the "ghost" number line to keep track of their points on the scoreboard. Further, in order to keep track of their opponents score, players have to convert between fractions and decimals.
Exactly–Flip In *Exactly–Flip,* we added a rule that requires players to switch between the fraction side and the decimal side anytime their teacher calls "Flip"	(A) fraction/decimal magnitude representation, (C) Conceptual fraction/decimal arithmetic, (D) conversion between fractions and decimals, and (E) fraction/decimal comparison In addition to skills practiced during *Exactly,* players also must engage in fraction/decimal conversions and comparisons several times during the game in order to keep track of their own and their opponents score.
Associated fraction/decimal skills target:(A) fraction/decimal magnitude representation, (B) reduce "whole number" bias, (C) Conceptual fraction/decimal arithmetic, (D) conversion between fractions and decimals, (E) fraction/decimal comparison, and (F) Automaticity.	

Table S2

Summary Statistics of Fraction Ball Moderators and Raw Scores of Outcomes for the Pilot Experiment

	Full Sample			Control			Treatment			
Variable	N	Mean	SD	N	Mean	SD	N	Mean	SD	*p*
Compliance	69	80%		37	78%		32	81%		.77
Male	69	55%		37	51%		32	59%		.51
5th grade	69	51%		37	51%		32	50%		.91
6th grade	69	49%		37	49%		32	50%		.91
Teacher Rating: Low	69	30%		37	27%		32	34%		.52
Teacher Rating: Average	69	30%		37	30%		32	31%		.89
Teacher Rating: High	69	39%		37	43%		32	34%		.46
Dosage in days							32	3.59		
Attrition from Pre-test to Post-test	69	1%		37	3%		32	0%		.36
Missing Pre-test	69	3%		37	5%		32	0%		.19
Missing Posttest	69	1%		37	3%		32	0%		.36
Missing 90% not-timed post-test items	69	4%		37	5%		32	3%		.65
Pretests										
Timed fraction to decimal conversion	67	17%	25%	35	17%	22%	32	18%	27%	.99
Timed decimal to fraction conversion	67	28%	27%	35	27%	25%	32	29%	30%	.77
Timed fraction addition	67	53%	29%	35	57%	27%	32	49%	31%	.27
Untimed Fraction to Decimal Conversion	67	34%	34%	35	34%	31%	32	35%	37%	.85
Untimed Decimal to Fraction Conversion	67	55%	36%	35	54%	35%	32	56%	38%	.76
Untimed Fraction and Decimal Addition	67	32%	35%	35	33%	34%	32	30%	35%	.71
PAE 0 to 1	67	11%	13%	35	9%	11%	32	14%	15%	.07
PAE 0 to 5	67	21%	15%	35	18%	14%	32	24%	16%	.11

Table S2 *(Continued)*

Post-tests										
Timed fraction to decimal conversion	68	21%	22%	36	17%	18%	32	27%	25%	.08
Timed decimal to fraction conversion	68	31%	27%	36	27%	24%	32	37%	29%	.20
Timed fraction addition	68	53%	26%	36	55%	28%	32	50%	25%	.47
Untimed Fraction to Decimal Conversion	68	51%	42%	36	49%	39%	32	54%	45%	.65
Untimed Decimal to Fraction Conversion	68	54%	41%	36	53%	38%	32	55%	44%	.79
Untimed Fraction and Decimal Addition	68	41%	39%	36	40%	40%	32	43%	39%	.76
PAE 0 to 1	68	13%	13%	36	11%	12%	32	15%	14%	.20
PAE 0 to 5	68	21%	14%	36	21%	15%	32	21%	13%	.92

Note.* p <0.05, ** p <0.01, *** p <0.001. p-value is for based on a two-tailed t-test comparing of each outcome by treatment group. Teacher ratings were collected from teachers based on their own criteria of math ability for each student in comparison to other students in the same class. Raw scores were standardized using average grade standard deviations in the control group. Indices are only shown as standardized scores to facilitate interpretation as indices contain raw scores and natural log transformed percent average error scores from the number line items. PAE = percent absolute error. The N in the Variable section refers to total sample possible, including student attrition.

Table S3

Summary Statistics of Fraction Ball Standardized Scores of Outcomes for the Pilot Experiment

	Full Sample			Control			Treatment			
Construct	*N*	Mean	SD	*N*	Mean	SD	*N*	Mean	SD	*p*
Pretests										
Average Score	67	-0.00	1.14	35	0.08	1.03	32	-0.09	1.25	0.54
Near Transfer	67	-0.00	1.11	35	0.08	1.06	32	-0.08	1.18	0.56
Far Transfer	67	0.00	1.24	35	0.10	1.04	32	-0.11	1.43	0.49
Post-tests										
Average Score	68	0.00	1.11	36	-0.05	1.03	32	0.06	1.21	0.69
Near Transfer	68	0.00	1.09	37	-0.07	1.03	32	0.06	1.16	0.60
Far Transfer	68	-0.00	1.21	37	-0.04	1.04	32	0.04	1.38	0.80

Note. * $p<0.05$, ** $p<0.01$, *** $p<0.001$. p-value is for based on a two-tailed t-test comparing of each outcome by treatment group Raw scores were standardized using average grade standard deviations in the control group.

Table S4

Estimated Treatment Effects for the Pilot Experiment

	Full sample with No Covariates				Full sample with Pretest & Grade Covariates				Compliant Sample with Covariates			
Outcome Variable	*b*	*(SE)*	*p*	*N*	*b*	*(SE)*	*p*	*N*	*b*	*(SE)*	*p*	*N*
Posttest Composite Score	.11	(.27)	.69	68	.18	(.12)	.15	66	.12	(.11)	.30	54
Composite Near Transfer Items	.14	(.27)	.60	68	.19	(.14)	.16	66	.11	(.13)	.39	54
Composite Far Transfer Items	.08	(.30)	.80	68	.17	(.16)	.28	66	.15	(.16)	.33	54
Timed Fractions to Decimal Conversion	.54	(.31)	.08	68	.49	(.18)	0.01**	66	.55	(.20)	0.01**	54
Timed Decimal to Fraction Conversion	.43	(.27)	.12	68	.30	(.17)	.08	66	.27	(.18)	.14	54
Timed Fraction Addition	-.18	(.25)	.47	68	-.06	(.18)	.75	66	-.14	(.19)	.48	54
Untimed Fractions to Decimals Conversion	.12	(.27)	.66	68	.01	(.17)	.95	66	.00	(.18)	.99	54
Untimed Decimal to Fraction Conversion	.07	(.26)	.79	68	-.02	(.18)	.89	66	-.09	(.17)	.59	54
Untimed Fraction and Decimal Addition	.07	(.24)	.76	68	.10	(.17)	.54	66	.10	(.19)	.61	54
Percent Absolute Error on 0-1 Number Line	-.23	(.25)	.36	68	-.06	(.19)	.74	66	-.18	(.21)	.39	54
Percent Absolute Error on 0-5 Number Line	-.15	(.23)	.52	68	.01	(.17)	.96	66	-.08	(.18)	.66	54

Note. Number line scores were reverse coded so that positive scores indicate better performance. Covariate is a composite z-score of child performance on the pretests. Standardized scores are reported to allow for comparison across measures. Standard errors are in parentheses under the regression coefficient. Based on our preregistered hypotheses, the outcomes are ordered from largest expected impacts at the top to the smallest expected impacts at the bottom.

Table S5

Estimated Treatment Effects for the Efficacy Experiment

	Full sample with No Covariates				Full sample with Pretest & Grade Covariates				Full Sample not missing 90% of post untimed test with Covariates			
Outcome Variable	*b*	*(SE)*	*p*	*N*	*b*	*(SE)*	*p*	*N*	*b*	*(SE)*	*p*	*N*
Posttest Composite Score	.38	(.19)	.05*	158	.44	(.11)	<.001***	155	.40	(.11)	<.001***	153
Composite Near Transfer Items	.42	(.19)	.03*	158	.46	(.12)	<.001***	155	.43	(.12)	<.001***	153
Composite Far Transfer Items	.18	(.20)	.37	158	.17	(.14)	.23	155	.12	(.14)	.37	153
Timed Fraction to Decimal Conversion	.57	(.19)	<.001**	158	.50	(.14)	<.001***	155	.48	(.14)	<.001***	153
Timed Decimal to Fraction Conversion	.20	(.19)	.29	158	.24	(.13)	.07	155	.22	(.13)	.11	153
Timed Fraction Addition	-.12	(.17)	.46	158	-.09	(.13)	.50	155	-.13	(.13)	.34	153
Untimed Fraction to Decimal Conversion	.65	(.19)	<.001***	158	.70	(.15)	<.001***	155	.69	(.15)	<.001***	153
Untimed Decimal to Fraction Conversion	.39	(.17)	.03*	158	.53	(.13)	<.001***	155	.51	(.13)	<.001***	153
Untimed Fraction and Decimal Addition	.21	(.19)	.27	158	.25	(.16)	.12	155	.24	(.16)	.15	153
PAE on 0-1 Number Line	.10	(.16)	.56	158	.18	(.13)	.19	155	.13	(.13)	.33	153
PAE on 0-5 Number Line	.20	(.17)	.26	158	.17	(.15)	.27	155	.14	(.15)	.35	153

Note. Number line scores were reverse coded so that positive scores indicate better performance. Covariate is a composite z-score of child performance on the pretests and grade level. Standardized scores are reported to allow for comparison across measures. Standard errors are in parentheses under the regression coefficient. Based on our preregistered hypotheses, the outcomes are ordered from largest expected impacts at the top to the smallest expected impacts at the bottom.

Table S6

Teacher Fidelity to Fraction Ball Games Script Pilot Experiment

Games	Total possible	5th grade	6th grade	Mean	SD
Duration in Minutes					
Rapid Fire	35-40	55	37.5	46.25	10.31
Make it Count	35-40	55	37.5	40.33	0.5
Exactly	35-40	33	33	33	3.65
Flip	35-40	39	33.5	36.25	3.4
Activities Completed					
Rapid Fire	88	99%	75%	82%	14%
Make it Count	7	100%	100%	100%	0%
Exactly	7	93%	57%	75%	22%
Flip	6	100%	100%	100%	0%

Note. Students participated in Fraction Ball games during their regularly scheduled physical education class thus 5th grade sessions were always in the morning and scheduled for 35 minutes, while 6th grade sessions were always in the afternoon and scheduled for 40 minutes. Session duration depended on teacher flexibility that day. The number of activities completed pertains to the number of key tasks delineated in the Fraction Ball script that the teachers completed such as “Introduction to Make it Count” or “Exactly round 1”.

Table S7

Teacher Fidelity to Fraction Ball Games Script Efficacy Experiment

Games	Total possible	4th grade	5th grade	6th grade	Mean	SD
Duration in Minutes						
Rapid Fire	35-60	35	52	28	38.33	12.34
Make it Count	35-60	–	34	25.3	29.65	6.15
Exactly	35-60	17.3	44.3	33.3	38.8	7.78
Rapid Fire Target	35-60	–	38	45	41.5	4.95
Make it Count Ghost	35-60	41	41.45	36	39.48	3.03
Exactly + Flip	35-60	35.3	44.15	–	39.73	6.26
Activities completed						
Rapid Fire	90	57%	80%	64%	67%	0.12
Make it Count	20	–	69%	53%	61%	0.11
Exactly	20	30%	76%	50%	52%	0.23
Rapid Fire Target	11	–	91%	100%	96%	0.06
Make it Count Ghost	12	50%	85%	67%	67%	0.18
Exactly + Flip	20	48%	69%	–	59%	0.15

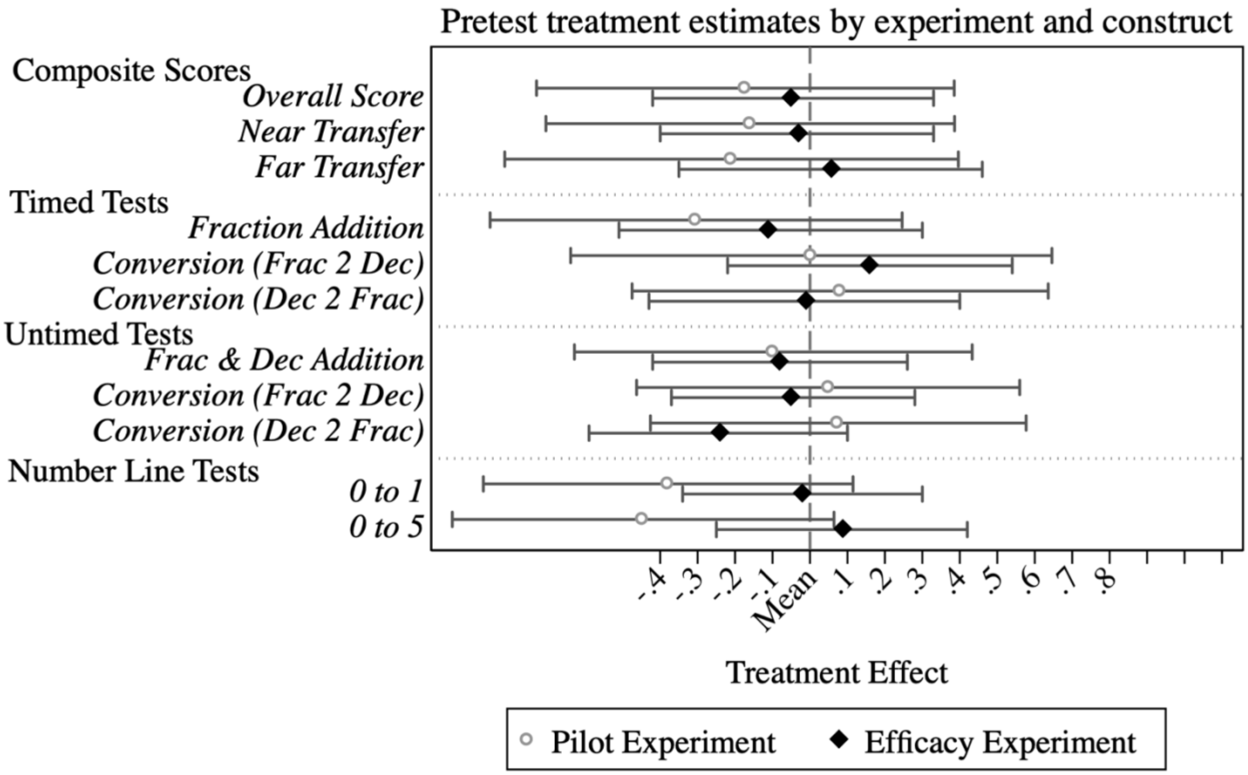
Note. Students participated in Fraction Ball games during their regularly scheduled physical education class which varied between the grades thus 4th and 5th grade sessions occurred in the morning and afternoon, however 6th grade sessions were always scheduled in the afternoon. The duration of the sessions also varied where 4th grade was able to conduct 45 to 60-minute sessions, 5th grade was able to conduct 50 to 60-minute sessions, and 6th grade was able to conduct 35 to 60-minute sessions. Session duration depended on teacher and basketball court availability. The table breaks down the average duration and activities completed by teachers within the same grade level, there were four teachers in 4th and 5th grade and two teachers in 6th grade. The number of activities completed pertains to the number of key tasks delineated in the Fraction Ball script that the teachers completed such as “Introduction to Make it Count” or “Exactly round 1”. (–) indicates missing observation data for the teachers within that grade level for the specific activity, researchers were unable to observe all sessions due to schedule conflicts.

Table S8

Estimated Treatment and Treatment by Moderator Effects on Overall Score

	Pilot Experiment				Efficacy Experiment			
Models	*b*	*(SE)*	*p*	*N*	*b*	*(SE)*	*p*	*N*
Grade Model								
Treatment	.16	(.15)	.30	66	.28	(.16)	.08	153
5th Grade X Treatment	-	-	-	-	.31	(.24)	.20	153
6th Grade X Treatment	-.01	(.23)	.96	66	-.06	(.28)	.84	153
Gender Model								
Treatment (Boys)	.32	(.16)	.04*	66	.43	(.15)	.006**	153
Gender X Treatment	-.35	(.21)	.10	66	-.08	(.22)	.71	153
Prior Knowledge Model								
Treatment (High Prior Knowledge)	.10	(.16)	.55	66	.41	(.13)	.002**	153
Prior Knowledge X Treatment	.12	(.21)	.58	66	-.02	(.22)	.91	153

Note. Betas represent the effect of treatment on overall score. The beta predictors are based on 3 separate models with overall composite score as criterion on treatment with controls for pretest score and grade + respective interaction term (treatment x grade; treatment x gender; treatment x prior knowledge). Standardized scores are reported to allow for comparisons across measures. Robust standard errors are in parentheses. In the Grade Model, 5th grade is the reference category for the Pilot Experiment and 4th grade is the reference category for the Pilot Experiment. * $p < .05$, ** $p < .01$



Note: Estimates represent regression coefficients for treatment status. They can be interpreted as treatment vs. control differences in grade SDs. Bars show 95% confidence intervals. *** $p < .001$, ** $p < .01$, * $p < .05$

Figure S1. Baseline estimates

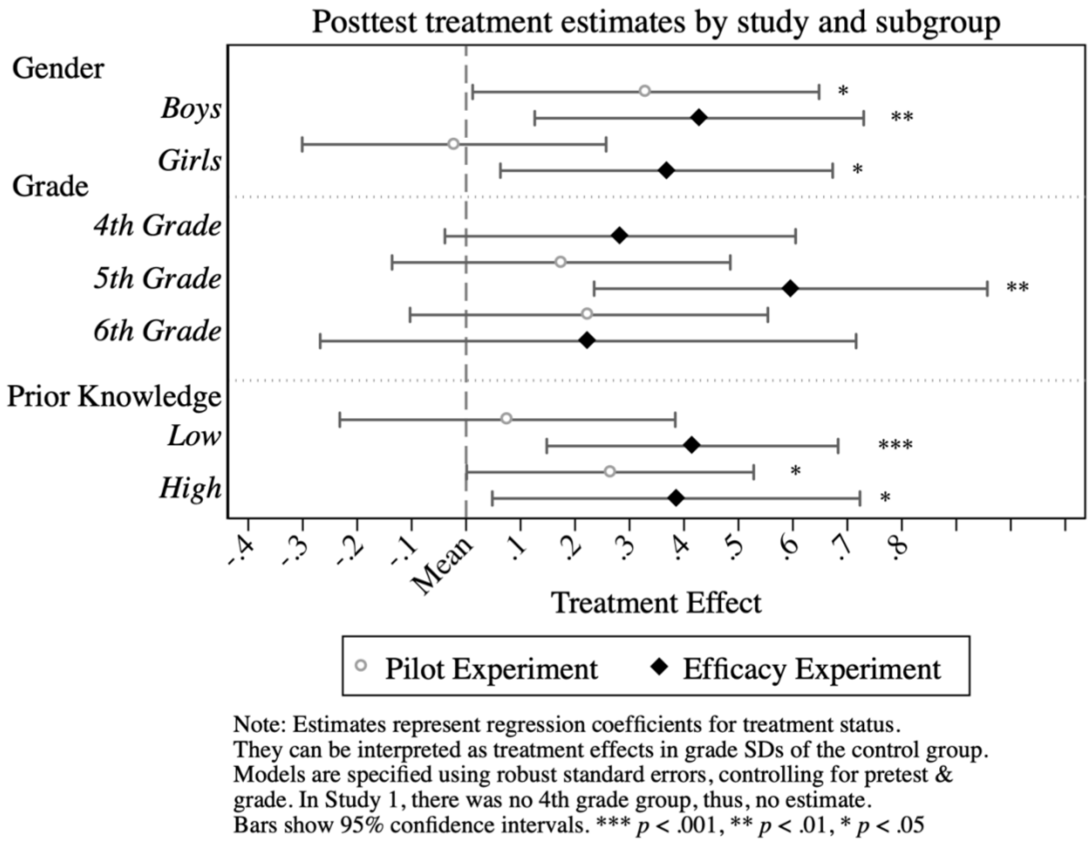


Figure S2. Posttest estimates of the impact of playing Fraction Ball on students' overall composite score of their rational number knowledge by subgroup.