Contemporary Educational Psychology 78 (2024) 102296

Contents lists available at ScienceDirect

Contemporary Educational Psychology

journal homepage: www.elsevier.com/locate/cedpsych

# Incorporating Design Based Implementation Research with a Randomized Controlled Trial to develop and evaluate the efficacy of playful rational number learning

Kreshnik N. Begolli [*], Vanessa N. Bermudez, LuEttaMae Lawrence [1], Lourdes M. Acevedo-Farag, Sabrina V. Valdez, Evelyn Santana, Daniela Alvarez-Vargas, June Ahn, Drew Bailey, Katherine Rhodes, Lindsey E. Richland, Andres S. Bustamante

*University of California, Irvine; School of Education, 401 E. Peltason Drive, Suite 3200, Irvine CA 92617, United States of America*

A R T I C L E I N F O

*Keywords:*
Rational numbers
Playful learning
Co-design
Randomized control trials
Math cognition

A B S T R A C T

We combine design-based implementation research with a pre-registered RCT to address a long-standing challenge in psychological science: How to use psychological principles to address real-world problems while designing and implementing interventions in the field. We posit this as a design methodology for optimizing the translation between psychological science and real-world applications. We tested the efficacy of an extensively co-designed version of a game-based rational number intervention, Fraction Ball, versus "business-as-usual" math instruction and physical education in a sample of 4th/5th grade Latine students ($N = 360$). Insights from nine co-design sessions with 20 teachers informed revisions and additions to a previous version of Fraction Ball, strengthening impacts across 10 of 12 rational number subtests. This methodology provides insights for translating psychological science research and scaling it to address real-world educational needs, such as play-based interventions that improve rational number understanding in authentic contexts.

## 1. Introduction

The kinds of scientific evidence provided by basic psychological research employing randomized control trial (RCT) approaches may not translate or scale effectively to create tools, programs, or initiatives that improve society (IJzerman et al., 2020; Lewis, 2021). Many theoretically informed interventions for improving students' learning and cognition have been found to be effective when delivered by researchers in controlled settings (Alibali & Knuth, 2018; Wenger, 1999), but have failed to replicate at scale (Lortie-Forgues & Inglis, 2019) or make inroads into educational settings (Pomerance et al., 2016). Current methodological approaches have been criticized for using non-representative convenience samples, lack of replication, stimuli that do not generalize to everyday settings, negligible effect sizes, and failures to account for the cultural, historical, political, and structural factors that are at play in translating scientific findings to real-world contexts (IJzerman et al., 2020).

In education, design-based implementation research (DBIR) approaches have become an increasingly popular methodology for addressing several of these challenges, largely overlooked in many program evaluations (Penuel et al., 2011). In contrast to many RCT studies in controlled settings that aim to develop generalizable knowledge, DBIR approaches are oriented towards designing and testing interventions that are adapted to local contexts and needs through a process of co-design with the users (in our case, teachers and students; Design-Based Research Collective, 2003). Co-design entails the creation of interventions by a highly-facilitated team of educators and researchers with predefined roles. The aim is to co-develop a prototype and qualitatively evaluate its import to address an educational need (Penuel et al., 2007). Yet, most DBIR studies do not utilize methods that

* Corresponding author.
*E-mail addresses:* kbegolli@uci.edu (K.N. Begolli), vnbermud@uci.edu (V.N. Bermudez), lu.lawrence@usu.edu (L. Lawrence), lfarag@uci.edu (L.M. Acevedo-Farag), svaldez@ebrooke.org (S.V. Valdez), santane2@uci.edu (E. Santana), dalvare5@uci.edu (D. Alvarez-Vargas), junea@uci.edu (J. Ahn), dhbailey@uci.edu (D. Bailey), ktrhodes@uci.edu (K. Rhodes), lerich@uci.edu (L.E. Richland), asbustam@uci.edu (A.S. Bustamante).
[1] Utah State University; Instructional Technology and Learning Sciences, 2830 Old Main Hill, EDUC Room 215, 953 East 700 North, Logan, UT 84322-2830, United States of America.

https://doi.org/10.1016/j.cedpsych.2024.102296

Available online 22 June 2024
0361-476X/Published by Elsevier Inc.

estimate causal effects of interventions on outcomes, hindering the generalizability of principles uncovered through typical DBIR (McCandlis et al., 2003).

The current study presents a two-step cyclical process that sheds light on how these different but complementary approaches may combine to create usable interventions tailored to local contexts and needs. This approach is grounded in research and rigorously evaluated to assess effectiveness. We report the results of a program of research combining DBIR and an RCT evaluation in a field experiment. The approach taken to intervention design and evaluation addresses several issues discussed by IJzerman et al. (2020). Specifically, our research includes historically underrepresented populations, incorporates everyday contexts and real-world stimuli, and responds to cultural and structural factors present in authentic educational settings. Furthermore, our approach evaluates the efficacy of the intervention we co-designed with educators, using a well-powered, internally valid RCT design implemented in public schools.

In this study, we employ a combination of DBIR and RCT methodologies to build on a single previous pilot and efficacy study of Fraction Ball–an outdoor play-based math intervention (referred to as pilot study hereafter). We aimed to apply insights from psychological science in complex environments where cognitive processes interact with dynamic social, cultural, and structural factors through the integration of teacher expertise and psychological science literature. In Fraction Ball, the traditional basketball court and rules have been reconstructed to support students' mental representations of rational numbers on a number line. In two pilot experiments, Fraction Ball generated positive impacts on students' rational number knowledge and conversions between fractions and decimals at one school site (see Fig. 1; Bustamante et al., 2022). However, the qualitative insights from teachers in the pilot study also raised concerns about Fraction Ball's potential for adoption, feasibility, and sustainability when scaling across multiple school sites and addressing problems of practice. For example, teachers expressed concerns about integrating the intervention into their curricula, aligning it with state standards in mathematics, and scheduling outdoor activities into their instructional time. Further, although previous studies were implemented in a school, it was a single school with a small number of teachers implementing the intervention, and a researcher was nearly always present. Finally, the control group in the original studies participated in business-as-usual physical education, which did not have a specific math focus. Thus, previous findings give some indication that when implemented with fidelity, the intervention is a useful supplemental mathematics activity. However, evidence for the usefulness of Fraction Ball incorporated into a curriculum would require an evaluation that accounts for on-the-ground challenges of a classroom and compares the intervention to existing PE and math curricula.

To contextualize our combined DBIR and RCT approach, we briefly review the affordances and constraints of DBIR and RCT. Then, we provide a brief literature review highlighting key learning principles that influenced our designs and their alignment with teacher expertise and practical experience.

## 2. Conceptual framework for combining RCT and DBIR

### 2.1. Randomized control Trials

Randomized Control Trials in education aim to generate estimates of the causal effects of interventions by randomly assigning participants (participant groups) to either a treatment group receiving the intervention or a control group receiving no treatment or an alternative treatment (e.g., Shadish, Cook, & Campbell, 2002). Because RCTs estimate causal relationships between an intervention and its impacts on student outcomes, it is broadly regarded as the "gold standard" in education research, heavily shaping U.S. education policies such as the No Child Left Behind Act (NCLB) and the Every Student Succeeds Act (ESSA, 2015; Welsh, 2021). The Institute of Education Sciences (IES) prioritizes RCT evidence in their Standards for Educational Excellence as reflected in the "levels of certification" by the What Works Clearinghouse–a U.S. federally funded authority on educational recommendations, policy, and interventions (Polikoff and Conaway, 2018; Welsh, 2021). However, it is infeasible to expect RCTs alone to guide us to optimal education policy and practices because designing educational interventions requires more decisions than can be feasibly evaluated through RCTs.

Design-based implementation research methods can facilitate the exploration, rapid testing, and feasibility of these decisions based on combined researcher and community expertise in order to develop an intervention prototype that would be most likely to be adopted, show success, and scale with relatively low costs compared to conducting an RCT. Randomized control trials can provide a relatively precise quantitative answer to a well-specified question; in multi-armed RCTs, a few questions can be asked simultaneously. However, designing educational practices invariably requires making a larger number of decisions than can realistically be evaluated in independent randomized treatment arms. Consider a set of 10 lessons, each designed to be delivered by a teacher to a classroom of students. Some education decision makers must decide how to sequence the lessons. However, because there are 10!, or approximately 3.6 million possible sequences of these 10 lessons, any evaluation of which sequence to prefer over others would require a decision about which sequences are under consideration. Of course, this task is only a microcosm of the problem of how to design an educational intervention (See Koedinger et al., 2013 for a related discussion). Therefore, RCTs should be complemented by a method for designing interventions, preferably one that is rigorous and systematic.

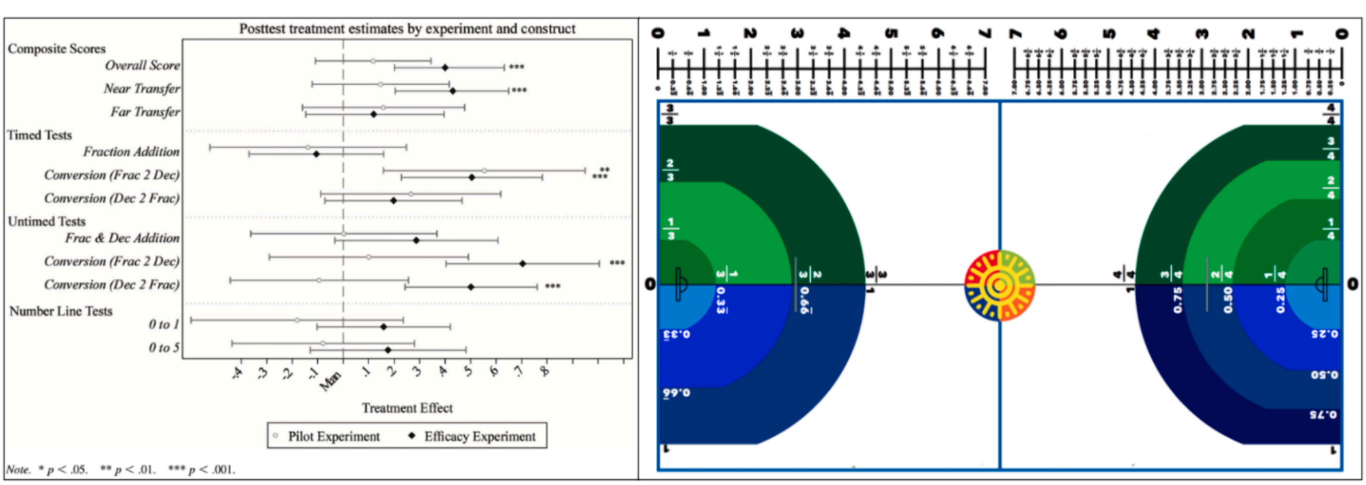



**Fig. 1.** Previous Study Posttest Treatment Estimates and Court Design.

### 2.2. Design based implementation research

To address this challenge, we draw from methodological approaches that center on-the-ground perspectives and goals of practice partners (in our case, teachers) through the process of co-design, which is at the core of DBIR (Penuel et al., 2011). This user-centered approach privileges users' subjective challenges and incorporates them into interventions, often by deliberating together with researchers around possible solutions. This process involves modifying interventions, which users then piloted and vetted. This step helps ascertain the viability of the intervention for implementation and identifies practical flaws before conducting costly, large-scale experiments. Co-design also affords the integration of users' key instructional needs and expertise in navigating organizational and logistical barriers around implementation within a system (Penuel et al., 2011). These obstacles account for the varied roles, needs, and schedules of people necessary for the success of operational and systemic goals at the school and district levels.

Penuel and colleagues (2011) summarized seven characteristics of co-design: 1) creating or improving a concrete innovation (e.g., adding classroom components to Fraction Ball); 2) understanding current practices and contexts (e.g., aligning Fraction Ball with math curriculum and standards); 3) allowing for flexibility in target outcomes (e.g., including analyses of distal measures such as state tests, which have high priority for teachers and districts); 4) creating events or processes that build understanding of the resource to catalyze the work (e.g., adding professional development materials, adjusting the curriculum to meet teacher capacity); 5) fitting co-design sessions within educators schedules; 6) highly-facilitated and well-defined roles of team members (e.g., empowering teacher expertise); 7) central accountability of the end product which rests with the lead investigator of the research team (and ensures continued alignment with evidence based practices). It is the role of the researchers in the co-design process to draw the themes and insights from users' experiences and design artifacts and integrate them with theoretical principles from psychological science (Penuel et al., 2011; Yeager et al., 2016).

Research conducted using design-based approaches has received criticism for not incorporating methods typically employed by RCTs (i.e., a control group and random assignment) to make causal claims and evaluate the impacts of designed learning environments on outcomes more traditionally assessed by RCTs, such as student learning (Design-Based Research Collective, 2003; McCandliss et al., 2003). Instead, typical DBIR studies frequently center around small-scale qualitative measures of success, such as recorded classroom interactions, student work, and other artifacts, formative and summative assessments, and other artifacts. The small-scale, rapid cycle measurement approach has resulted in rich descriptive data of the local contexts and conditions that enable data points on which quick iterations and refinements of instructional designs can be made (Penuel et al., 2011). More broadly, rich descriptions and insights gained from DBIR can reveal important behavioral phenomena as they manifest in everyday complex contexts, such as school systems, which provides opportunities to develop conjectures and theorize about scientific principles of learning processes (McCandlis et al., 2003). However, schools and educators value student learning and seek strong evidence of student learning outcomes before adopting a new program (Honig & Coburn, 2008; Slavin & Fashola, 1998). Therefore, causal impacts on student learning are an important component for scaling, dissemination, and ultimately, the impact of a program (Honig & Coburn, 2008).

Most published DBIR studies report improvements in learning environments. As noted above, it is not realistic to expect a well-powered RCT to evaluate the efficacy of every feature considered during the design process. However, because in this case, Fraction Ball was intended to improve student learning outcomes in our population of interest (relative to not participating in Fraction Ball), an RCT was a useful complement to the DBIR process for determining whether Fraction Ball reached this goal.

### 2.3. The common ground between RCT and DBIR

Taken together, RCT and DBIR approaches are undergirded by epistemological stances that have been traditionally in opposition with one another in terms of their aims. RCT approaches have been primarily employed to make causal inferences about the efficacy of an intervention. On the other hand, DBIR approaches strive to characterize the contextual conditions in order to iteratively improve an instructional design. Of course, both approaches can also lead to conjectures and theory refinement about learning processes.

In Fraction Ball, we harness the strengths of DBIR and RCT approaches in a two-step cyclical process. First, through DBIR, we co-design interventions that integrate psychological science principles and attend to on-the-ground challenges of local school contexts to iterate on our designs through rich descriptions of contextual factors and student–teacher interactions. Our research team draws themes from a series of co-design sessions, identifies theory and principles that align with those themes, which we use to modify the intervention, vetted by our teachers and students. Further, results from our prior pilot RCT evaluations demonstrated positive impacts, but limited to a few domains, which oriented the team towards improving aspects of the intervention where effects around rational number concepts were mixed and/or not impactful. As such, our co-design discussions also incorporate topics deriving from prior RCT data. Then, we utilize an RCT to assess how the intervention, informed by co-design and the prior RCT evaluation, affects student outcomes, enabling us to make a causal inference about its impact on student performance. The two-step cycle may be repeated to further improve intervention design and evaluate its effects.

This approach embraces the complementary purposes served by these traditionally separate epistemologies (Begolli & Richland, 2017; Brown, 1992, p. 199; McCandliss et al., 2003). It provides a framework for a two-way exchange of rich qualitative data and quantitative measures that allow for conjecture building of phenomena that can be tested in controlled and/or school settings. In summary, we use DBIR and RCT approaches to iteratively improve interventions grounded in psychological science, attend to the challenges of the local context, identify gaps in intervention design, and test causal hypotheses about their efficacy.

Given that our intervention draws heavily from psychological science literature, next, we review how theory and principles from this discipline have informed the Fraction Ball design and how the two-way exchange that occurred between researchers and teachers influenced the intervention design.

## 3. Integrating psychological science and teacher expertise through Co-Design

The design of the Fraction Ball court draws heavily from research in mathematical cognition, particularly the Integrated Theory of the Number Line, positing that a mental representation of the number line is key for integrating students' conceptual understanding of whole and rational number magnitudes (Siegler & Lortie-Forgues, 2014). Empirical evidence from several game intervention studies with the number line as a core component further supports that number line games can enhance students' comprehension of fraction magnitudes (Fazio et al., 2016). Intensive interventions that use visualizations of the number line and magnitude understanding as core components influence both fraction magnitude and arithmetic knowledge with students struggling with fraction concepts and students with disabilities (Barbieri et al., 2020; Bottge et al., 2014; Fuchs et al., 2013, 2016; Schumacher et al., 2018). In contrast, area models seem to be less effective than number line models, as evidenced by data from 2nd and 3rd graders (Hamdan & Gunderson, 2017).

The Integrated Theory of the Number line informed the layout of the court and point system, with greater distances from the court corresponding to higher point values (Fig. 1). The Fraction Ball scoreboard–a large number line on the side of the court–serves to reinforce visualizations of fraction and decimal magnitudes, such that students move further along the number line as they gain more points. These visual aids help students spatially discern rational numbers and gain familiarity with whole number magnitudes, and in turn, possibly facilitate a deeper grasp of the part/whole relationship between numerators and denominators representing equal parts of a whole (Seethaler, Fuchs, Star, & Bryant, 2011; Siegler, Thompson, & a., & Schneider, M., 2011; Vamvakoussi & Vosniadou, 2010).

The original version of Fraction Ball occurred exclusively outdoors on the basketball court using the number line. However, co-design teachers from our pilot and current studies were concerned about meeting their curricular goals and state standards. They stressed the necessity of drawing relationships between procedural knowledge required to carry out arithmetic operations in traditional classroom settings and the concrete number line representations on the Fraction Ball court. This input strongly aligned with prior research suggesting that procedural and conceptual knowledge work cyclically, each reinforcing the other (Rittle-Johnson et al., 2015). It also aligns with analogical reasoning theories, where drawing relations, aligning, and mapping between core elements of two or more representations leads to schema building (Gentner, 1983; Gick & Holyoak, 1983). For example, on the court, students can "see" that 1/4 is equal to 0.25 and also notice that 1 is bigger than both despite having fewer digits (Begolli & Richland, 2016). To relate these concepts with addition procedures, co-design teachers developed in-classroom activities where fraction and decimal arithmetic sentences (e.g., 1/4 + 3/4 and 0.25 + 0.75) are superimposed on top of a number line. In this manner, students can also "see" (align) that adding fractions and their decimal counterparts results in the same magnitude despite having different (misaligned) symbolic notations.

Failing to notice misalignments, particularly around fraction/decimal magnitudes, can lead to common misconceptions around rational number concepts. "Whole number bias" characterizes misconceptions of both rational number notations such that "more" or "larger" numbers are thought to have bigger magnitudes (e.g., considering a/5 larger than a/2 or 0.10 larger than 0.1; Hiebert & Wearne, 1983; Tian & Siegler, 2017). Additionally, procedures vary between notations and are often confused when taught as a sequence of steps that are memorized and followed in order to get to the correct answer (Richland et al., 2012). For example, students often erroneously add both the numerators and the denominators of fractions (e.g., 1/5 + 1/5 = 2/10) or arrive at the wrong answer for decimals with a different number of digits after the decimal place (e.g., 3.7 + 1.42; 12 % of 6th graders were correct; Hiebert & Wearne, 1983). Highlighting the alignments and misalignments between notations and procedures of fractions and decimals could help overcome common misconceptions as students can build on the more familiar notation (Wang & Siegler, 2013). However, the cognitive process of drawing relationships is challenging and can burden children's cognitive resources (Richland et al., 2017).

To reduce cognitive load, we limited the number of fraction representations included in the court design and the intervention to provide opportunities to build a more robust conceptual understanding around more commonly used magnitudes, such as fourths and thirds. Teachers validated the use of these benchmark fractions (thirds and fourths) through our co-design, which also aligns with fraction representations recommended in the Common Core state standards for mathematics (Common Core State Standards Initiative, 2010).

From these insights, we co-developed a new set of classroom activities that complement the outdoor games. Number line representations became a substantial component of classroom activities that emphasized number line segmentation (e.g., segment a 0–1 or 0–5 number line into fourths), magnitude estimation (placing 1/4 and 0.25 on the number line), and addition (e.g., superimposing and aligning 1/4 + 3/4 = 1 on top of the number line; see the Teacher Activity Guide in Appendix B for examples).

## 4. The current study

Here we report findings from the co-design process and an experimental evaluation of a revised version of Fraction Ball that integrated lessons learned from our pilot RCT, insights from nine co-design sessions with educators, and learning principles from psychological science. Through the two-step DBIR and RCT approach, first, we co-identified and co-developed with teachers key differences in the intervention, importantly, the addition of six in-classroom activities that integrate with the six outdoor games and replaced regular math instruction (see Method section for more detail). In our second step, utilizing an RCT, these elaborations on the intervention enabled us to evaluate the intervention as a substitute for business-as-usual math instruction rather than a supplementary activity during PE time. This design allowed us to draw conclusions about the efficacy of Fraction Ball when implemented as an integrated part of a school's math curriculum.

Co-designed developments and adaptations to Fraction Ball demonstrate the promise of combining DBIR and RCTs for leveraging knowledge from psychological science, teacher practical knowledge, and prior evaluations to create usable interventions and generate strong causal evidence of student learning utilizing experimental designs. This study provides an important entry point for innovative psychological approaches that target well-documented public and social needs, such as improving school curricula and helping students reach state standards in mathematics education. We hypothesized that the revised, co-designed Fraction Ball intervention would strengthen students' mental representations of rational number magnitudes and other subskills compared to students who received business-as-usual math instruction and participated in PE. The study's research questions, hypotheses, and analyses were pre-registered at https://osf.io/kjqmz prior to posttest data collection.

### *4.1. Research question & hypothesis*

We pose two main research questions in this study:

What were the key adaptations that altered the Fraction Ball intervention through co-design with teachers?

Does the Fraction Ball intervention as implemented by teachers improve students' rational number reasoning skills relative to "business-as-usual" math instruction and PE?

We also pose a three part exploratory research question:

Are the impacts of Fraction Ball on students' learning moderated by

A. grade,
B. sex, or
C. prior knowledge?

Based on the previous versions of Fraction Ball (Bustamante et al., 2022), we hypothesized that students in the Fraction Ball intervention group would improve their scores on the rational number reasoning battery. We pre-registered a conservative estimate of about 0.30 to 0.44 SDs, but also expect to strengthen these impacts and improve student knowledge on other components of rational number understanding previously not found as a result of iteratively co-designing with teachers to improve the intervention. At the same time, the current evaluation differed in ways that could produce smaller effect sizes. For example, the current evaluation included an in-class component, which would hopefully increase learning. However, as a result, the current evaluation replaced six regular math lessons with Fraction Ball lessons, so the treatment–control contrast could also plausibly be *weaker* than in previous implementations.

Based on previous results, we also hypothesized that the impact of the intervention would be larger on near versus far transfer items and larger on conversions between fractions and decimals than other kinds of problems. However, as a result of changes that arose from co-designing with teachers, we hoped to elicit statistically significant impacts on number line estimation, adding fractions with decimals, and timed fraction addition; three types of problems for which we did not observe impacts in previous evaluations. Finally, we explored whether the impacts of the intervention on students' learning was moderated by grade, sex, or prior knowledge (based on previous findings, we did not anticipate substantial heterogeneity by these factors). We included these exploratory factors given the abundance of literature suggesting implications of prior knowledge on learning (Simonsmeier et al., 2022). We included gender as a factor of interest, both to monitor heterogeneity in the impacts of Fraction Ball across socially important demographics and because gender differences in math achievement are present in U.S. math assessments (generally favoring boys in districts serving higher-income students and girls in districts serving lower-income students; Reardon et al., 2019). Analyses of pre-registered research questions on students' enjoyment of the Fraction Ball activities and self-reported emotions towards the math test are reported elsewhere.

## 5. Method

First, we present our diversity, equity, and inclusion (DEI) strategy as it relates to our co-design process, then we describe the co-design and RCT evaluation methodologies.

Our project team has advised and collaborated with our funder through the EF + Math Program, to develop guidelines around inclusive research and development to develop an adaptable executive function and math (EF + Math) learning platform that promotes equity through co-design methodologies. We work directly and actively with a team of partner educators, and their design ideas are directly adapted into new aspects of our intervention. Active involvement of educators from our partner district fosters a collaborative environment, enriching the design process. Equity considerations extend to co-designed technological components, catering to diverse learning needs while prioritizing scalability and technological value. Our approach embraces flexible meeting modalities, reflecting a commitment to diversity, equity, and inclusion, and addressing biases while valuing diverse perspectives. Tangible benchmarks involve diverse co-design groups, systematically documenting emergent themes, including aspects of culture, gender, ableism, and inclusion that influence design. Feedback refines our prototype, aligned with standards and real-world needs. Inclusive research and development, driven by human-centered methodologies, involve educators, synthesizing community insights into solutions that reflect context and meaning.

### *5.1. Co-design methods and analysis process*

Two schools participated in the co-design process, including 24, 3rd through 6th-grade teachers (five PE [three female, two male], four math [one female, three male], one special education [female], and 14 all subject teachers [10 female, four male]), a female curriculum specialist, and two principals (one female, one male). Most teachers had either played Fraction Ball with their students or played Fraction Ball throughout our co-design work together. We held five 90-minute sessions via Zoom and in person; two sessions were repeated to accommodate teachers' schedules. Based on teacher feedback from our pilot study, we developed broad goals to create classroom activities and adapt the games on the court. Through an iterative co-design process our sessions focused on: a) introducing the games to teachers, b) play-testing an existing game, c) generating lesson plans and activity ideas to make connections between Fraction Ball games, common core state standards in mathematics, and the district curriculum for 4th and 5th grade, and d) teacher gallery walks of new classroom/court activities to iterate and refine ideas into cohesive lesson plans.

We collected video data, observation notes, design artifacts, and reflections to revise and develop intervention materials that were brought back to teachers for feedback and iteration. To analyze the key adaptations that emerged during the co-design, we conducted affinity diagramming, a grounded theory, thematic analytic technique used in design fields to synthesize design ideas (Hanington & Martin, 2019). First, two researchers examined all data and identified ideas and feedback to identify individual data points. We collected 305 data points from the co-design sessions, which were iteratively clustered into themes. To achieve reliability, two researchers analyzed the co-design data, negotiating themes and reviewing the data until consensus was reached. Our thematic analyses highlight the three most significant adaptations made to the Fraction Ball intervention using design vignettes. We bound these vignettes based on our "intrinsic interest" (Merriam, 1998, p. 28) in identifying key adaptations that teachers valued most to improve Fraction Ball's usability, feasibility, and adoptability.

### *5.2. RCT participants*

Study procedures were approved by the University's IRB. Co-design teachers and schools did not participate in the evaluation of the study to avoid spillover effects. Participants in the evaluation of the study were 16 female teachers (10, 4th-grade; five, 5th-grade; one, 4th/5th-mixed-grade teacher) and 360 students (218, 4th-grade; 142, 5th-grade; 52% female, 48% male; 47 % English Language Learners (ELL); 97 % Hispanic or Latino, 83 % free or reduced lunch, and 10 % student disability) from four non-co-design schools in Southern California. The sample demographics resembled those of the district as a whole (45 % ELL, 96 % Hispanic or Latino, and 81 % free or reduced lunch; Santa Ana Unified School District, 2022). Classroom size ranged from 16 to 45 students ($M$ = 23). Randomization was matched based on pretest scores aggregated at the classroom level (treatment, $n_{\text{teachers}} = 8$, $n_{\text{students}} = 198$; control, $n_{\text{teachers}} = 8$, $n_{\text{students}} = 162$; Table 1; Table A1). Teachers and students were recruited via email sent to the school principal and were unaware whether they would be assigned to treatment or control until after the pretest. Treatment teachers were provided a stipend of $250, and control teachers with $50 for classroom materials.

To assign teachers/classrooms to the treatment condition, students' pretest scores were aggregated at the teacher/classroom level. Given that two classrooms scored noticeably higher than the other classrooms at pretest, classrooms were blocked into eight pairs based on pretest scores (e.g., the first block consisted of the two classrooms with the highest scores, the second block consisted of the third and fourth classrooms with the highest scores). Classrooms were randomly assigned within each block, such that one classroom per block was assigned to the Fraction Ball intervention during math instruction and PE ($n_{\text{teachers}} = 8$, $n_{\text{students}} = 198$) and the other was assigned to the control group, math instruction and PE as usual ($n_{\text{teachers}} = 8$, $n_{\text{students}} = 162$). As a result, each school had at least one teacher in the intervention group and one in the control group. Descriptive statistics are provided for moderators and student-level measures split by treatment group in Table 1 and by treatment group and grade level in Table A1.

In the absence of direct classroom observations within control groups, our interpretation gains depth through a comprehensive examination of curriculum plans outlined in the provided curriculum maps. Control classrooms utilized district-endorsed curricula, Math Expressions, and the Irvine Math Project. Curricula were supplemented by Mathematical Assessment Resource Service (MARS) tasks, ST Math technology, and the Problem of the Month.

During the evaluation window 4th grade teachers were expected to focus on Collecting Data and Angle Measurements. This focus followed prior coverage of Fraction Equivalence and Decimals concepts and other rational number concepts, constituting approximately one-third of the overall curriculum. In parallel, teachers in 5th grade classrooms were expected to focus on two-dimensional shapes and volume after having extensively addressed rational number concepts in approximately two-thirds of the curriculum.

We used the PowerUpR package (Bulus et al., 2021) in R to conduct a power analysis. Our sample contained 16 teachers with approximately

**Table 1**
Descriptive Statistics for Demographic Variables and Raw Scores of Outcomes.

| | Full Sample | | | Control | | | Treatment | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | N | Mean | SD | N | Mean | SD | N | Mean | SD | *b* | *p* |
| 4th grade | 360 | 61 % | | 162 | 59 % | | 198 | 62 % | | 0.03 | 0.89 |
| 5th grade | 360 | 39 % | | 162 | 41 % | | 198 | 38 % | | −0.03 | 0.89 |
| Female | 333 | 52 % | | 148 | 45 % | | 185 | 58 % | | 0.13 | 0<.01** |
| Hispanic or Latino | 333 | 97 % | | 148 | 98 % | | 185 | 96 % | | −0.02 | 0.44 |
| English language learner | 333 | 47 % | | 148 | 50 % | | 185 | 44 % | | −0.06 | 0.65 |
| Free or reduced lunch | 325 | 83 % | | 146 | 84 % | | 179 | 82 % | | −0.01 | 0.88 |
| Student disability | 333 | 10 % | | 148 | 9 % | | 185 | 10 % | | 0.01 | 0.85 |
| Attrition from pre-test to post-test | 342 | 8 % | | 151 | 10 % | | 191 | 6 % | | −0.04 | 0.17 |
| Missing pre-test | 360 | 5 % | | 162 | 7 % | | 198 | 4 % | | −0.03 | 0.25 |
| Missing post-test | 360 | 7 % | | 162 | 9 % | | 198 | 6 % | | −0.04 | 0.19 |
| ***Pre-tests*** | | | | | | | | | | | |
| Missing 90 % not-timed pre-test items | 342 | 0.3 % | | 151 | 0.7 % | | 191 | 0 % | | −0.01 | 0.30 |
| Timed fraction to decimal conversion | 342 | 10 % | 22 % | 151 | 14 % | 27 % | 191 | 7 % | 17 % | −7.36 | 0.39 |
| Timed decimal to fraction conversion | 342 | 20 % | 24 % | 151 | 20 % | 22 % | 191 | 19 % | 25 % | −1.01 | 0.90 |
| Timed fraction addition | 342 | 48 % | 29 % | 151 | 48 % | 29 % | 191 | 47 % | 29 % | −0.50 | 0.94 |
| Untimed fraction to decimal conversion | 342 | 9 % | 21 % | 151 | 12 % | 25 % | 191 | 6 % | 15 % | −6.78 | 0.32 |
| Untimed decimal to fraction conversion | 342 | 39 % | 45 % | 151 | 36 % | 43 % | 191 | 41 % | 47 % | 4.41 | 0.77 |
| Untimed fraction and decimal addition | 342 | 25 % | 37 % | 151 | 33 % | 42 % | 191 | 18 % | 32 % | −14.79 | 0.25 |
| PAE fraction number line 0 to 1 | 342 | 21 % | 18 % | 151 | 22 % | 18 % | 188 | 21 % | 17 % | −1.81 | 0.67 |
| PAE fraction number line 0 to 5 | 342 | 28 % | 16 % | 151 | 26 % | 15 % | 188 | 29 % | 16 % | 3.27 | 0.43 |
| ***Post-tests*** | | | | | | | | | | | |
| Missing 90 % not-timed post-test items | 334 | 0.9 % | | 147 | 1.4 % | | 187 | 6 % | 5.3 % | −0.01 | 0.44 |
| Timed fraction to decimal conversion | 334 | 21 % | 32 % | 147 | 19 % | 33 % | 187 | 22 % | 32 % | 2.52 | 0.84 |
| Timed decimal to fraction conversion | 334 | 28 % | 30 % | 147 | 27 % | 32 % | 187 | 29 % | 28 % | 2.12 | 0.84 |
| Timed fraction addition | 334 | 56 % | 27 % | 147 | 56 % | 29 % | 187 | 55 % | 26 % | −0.91 | 0.89 |
| Untimed fraction to decimal conversion | 334 | 33 % | 40 % | 147 | 23 % | 37 % | 187 | 41 % | 41 % | 18.68 | 0.16 |
| Untimed decimal to fraction conversion | 334 | 48 % | 46 % | 147 | 44 % | 47 % | 187 | 50 % | 45 % | 5.87 | 0.64 |
| Untimed fraction and decimal addition | 334 | 39 % | 41 % | 147 | 34 % | 41 % | 187 | 44 % | 41 % | 9.91 | 0.45 |
| PAE fraction number line 0 to 1 | 324 | 17 % | 17 % | 141 | 20 % | 17 % | 183 | 15 % | 16 % | −4.86 | 0.22 |
| PAE fraction number line 0 to 5 | 324 | 23 % | 15 % | 141 | 24 % | 16 % | 183 | 22 % | 14 % | −1.79 | 0.69 |

*Note.* p-value is based on a two-tailed *t*-test comparing treatment and control, clustering standard errors by teacher. PAE = percent absolute error. The N in the Variable section refers to the total sample possible, including student attrition. ** $p < 0.01$.

23 students per class. We assumed an intraclass correlation coefficient (ICC) of 0.1. This is based on previous studies reporting on 4th and 5th graders math tests, suggesting an ICC range from 0.02 –.19 across three studies with sample sizes ranging from 54 to 1,284 students (Barbieri et al., 2020; Fuchs et al., 2019; Schoen et al., 2018). We assumed a small ICC because we were informed that classes were heterogeneously grouped. Our previous evaluation of Fraction Ball revealed a correlation of $r = 0.81$ between the composite pretest and posttest scores on a nearly identical test. Thus, we assumed that pretest scores would explain 64 % of the variation in posttest scores. Under these assumptions, our study design had 80 % power to detect an effect size of 0.34 SD, on the lower side of the range of predicted impacts.

### 5.3. *Procedure and fidelity observations*

Students completed the pretest two to three weeks prior to intervention and posttest within one-week of the end of the intervention. Intervention teachers were given lesson materials and attended a 90-minute professional development meeting prior to the intervention. Implementation ranged three-six weeks ($M = 3.75$). Trained members of the research team observed 25 % ($n = 22$) of the 88 total lessons completed by the teachers. Table 2 shows descriptive statistics by teacher for the implementation of the intervention, lessons completed, and duration and activities completed based on fidelity observations.

**Table 2**
Descriptive Statistics of Fraction Ball Implementation, Lessons Completed, and Fidelity Observations.

| | Implementation | | | Lessons Completed | | | Fidelity Observations | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Duration (Mins.) | | Activities Completed (%) | |
| Teacher | Weeks | Days | Teaching Sessions | Total | Classroom | Court Games | N | Classroom Lessons | Court Games | Classroom Lessons | Court Games |
| A | 4 | 5 | 8 | 8 | 4 | 4 | 0 | – | – | – | – |
| B | 3 | 11 | 13 | 12 | 6 | 6 | 4 | 27 | 44 | 73 % | 68 % |
| C | 3 | 9 | 15 | 12 | 6 | 6 | 6 | – | 44 | 67 % | 66 % |
| D | 6 | 4 | 8 | 8 | 4 | 4 | 0 | – | – | – | – |
| E | 4 | 7 | 18 | 12 | 6 | 6 | 1 | 30 | – | 41 % | – |
| F | 3 | 8 | 13 | 12 | 6 | 6 | 5 | 54 | 18 | 66 % | 37 % |
| G | 3 | 9 | 13 | 12 | 6 | 6 | 2 | – | 37 | – | 82 % |
| H | 4 | 7 | 12 | 12 | 6 | 6 | 4 | 24 | 35 | 86 % | 49 % |
| Mean | 3.75 | 7.5 | 12.5 | 11 | 5.5 | 5.5 | | 33.75 | 35.6 | 66 % | 60 % |
| SD | 1.04 | 2.27 | 3.34 | 1.85 | 0.93 | 0.93 | | 13.72 | 10.64 | 16 % | 17 % |

*Note.* Teaching sessions represent the number of blocked time teachers used to teach the classroom lessons and facilitate the court games. The intervention consisted of 12 lessons, out of which six were to be completed in the classroom and six in the court. Our fidelity of implementation thresholds was set at 20% of the intervention across all participating teachers and schools. However, due to scheduling conflicts between our team members and the schools, there was variation in observations between teachers.

## 6. Measures

### 6.1. Rational number skills

Students completed a 44-item assessment with a timed (28 items) section that assessed efficiency in rational number processing and an untimed (16 items) section that assessed rational number conceptual and procedural knowledge and magnitude understanding. Items were selected from the previous version of the Fraction Ball intervention (Bustamante et al., 2022) based on researcher-developed pre-registered item analyses. The assessment included eight mutually exclusive subtests to provide insight into the types of rational number skills impacted by the intervention. Items across all subtests were divided into near (24 items) or far transfer items (20 items). Near transfer items measured proximal knowledge of numbers and procedures to which students were directly exposed during the intervention containing thirds, fourths, and their decimal counterparts less than one, except for two items, which were 4 1/4 and 1.25. Far transfer items measured distal knowledge of numbers and procedures not directly presented during the intervention. Table 3 shows the assessment items by subtest and type of transfer item. These subtests reflect a combination of items previously used in the literature suggesting to capture mental representations of students' rational number concepts (Carvalho & da Ponte, 2017; Schneider & Siegler, 2010; Wang & Siegler, 2013).

**Timed Subtests.** Twenty-eight items assessed students' ability to quickly convert between fractions and decimals and add fractions within a three-minute period. These items comprised the three timed subtests: *timed fraction to decimal conversion* (eight items), *timed decimal to fraction conversion* (10 items), and *timed fraction addition* (10 items; same and different denominators, mixed fractions, and improper fractions). Items from each of the timed subtests were presented in an interleaved sequence. Internal consistency for the timed test was high at pretest ($\alpha = 0.90$) and posttest ($\alpha = 0.94$).

**Untimed Subtests.** Sixteen items assessed students' rational number conceptual and procedural knowledge. These items comprised the five untimed subtests: *fraction number line 0 to 1,* where students estimated the position of rational numbers on a 0 to 1 number line (four items), *fraction number line 0 to 5* (four items), *fraction to decimal conversion* (three items), *decimal to fraction conversion* (three items), and *fraction and decimal addition* (two items, e.g., 2/4 + 0.25). Internal consistency for the untimed test was high at pretest ($\alpha = 0.83$) and posttest ($\alpha = 0.90$).

**Table 3**
Near and Transfer Items by Subtest.

| Subtest | Near Transfer Items | Far Transfer Items |
|---|---|---|
| Timed fraction to decimal conversion | 1/4, 2/3, 1/3, 3/4, 1/2 | 6/3, 2/10, 12/3 |
| Timed decimal to fraction conversion | 0.33, 0.66, 0.50, 1.00, 0.75 | 0.40, 2.00, 0.80, 2.33, 4.50 |
| Timed fraction addition | 1/4 + 2/4, 2/3 + 1/3 | 1/5 + 3/5, 1/5 + 3/10, 1/4 + 1/2, 1 1/5 + 3 1/5, 2 3/4 + 1 3/4, 8/3 + 1/3, 4/10 + 5/10, 5/3 + 5/3 |
| Untimed fraction to decimal conversion | 1/3, 5/3, 1/2 | – |
| Untimed decimal to fraction conversion | 0.25, 1.25, 0.75 | – |
| Untimed fraction and decimal addition | 2/4 + 0.25 | 1/10 + 0.5 |
| Fraction number line 0 to 1 | 1/4, 1/4 + 2/4 | 4/5, 1/10 |
| Fraction number Line 0 to 5 | 4 1/4, 1/4 + 3/4, 1/3 | 12/4 |

*Note.* Untimed fraction to decimal and decimal to fraction conversion subtests did not have far transfer items.

### 6.2. Scoring of items

Items for the fraction number lines were scored using percent absolute error (PAE; Booth & Siegler, 2006), defined as PAE = [absolute value of (student's response – correct response)]/ (size of the number line). For items left blank, we imputed the missing value as incorrect in alignment with how teachers would typically score students, under the reasonable assumption that items left blank are far less likely to have been answered correctly than items attempted. For number line items, blanks were imputed with the 90th percentile PAE value for the item given our treatment of missing values as incorrect. Due to the nature of these items, instead of scoring missing values as zero (representing no error), we imputed them with a large PAE (90th percentile) to indicate low performance. The PAEs imputed were specific to each item because the maximum possible PAE depends on the distance from the furthest endpoint. PAE values were then transformed to a natural log scale and reverse-coded for higher values to represent better accuracy (Bustamante et al., 2022). Items for the other subtests were scored as correct or incorrect, with blanks imputed as incorrect. Fraction answers did not have to be in their simplest form to be scored as correct, and decimal answers had to be the exact values to the hundredths decimal place (no rounding) to be scored as correct. Items not attempted were scored as incorrect.

### 6.3. Composites

Average raw scores were calculated for each of the eight subtests at pretest and posttest (see Table 1). The average raw scores were standardized using a pooled standard deviation within each grade level of 4th and 5th grade using the entire sample at pretest and only the control group at posttest. An average standardized composite was calculated by averaging standardized scores across the eight subtests (Table A2). The same procedure was used with near- and far-transfer items to create near- and far-transfer composites (Table A2).

### 6.4. Data analysis

Multiple approaches were pre-registered to investigate the robustness of the treatment effects. Specifically, we investigated impacts on the overall, near transfer and far transfer composites, and each of the subtests. Our preferred specifications were regression models with standard errors clustered at the teacher level (Table 4) with overall math composite at pretest and grade level as covariates ($n = 316$), excluding students who were absent at pretest ($n = 18$) or posttest ($n = 26$). We also report estimates for a model with no covariates ($n = 334$) to test the sensitivity of estimates to the inclusion of baseline covariates. As robustness checks, we estimated models including covariates and excluding students who had missing data on 90 % or more of the untimed posttest items ($n = 314$; 2 students did not meet criteria). Other pre-registered specifications were generalized linear mixed (GLM) models using random intercepts at the teacher level (Table A3 and Table A4) and regression models using full information maximum likelihood (FIML) to address missing data and estimate treatment effects for the full sample ($n = 360$; Table A5). Finally, we conducted pre-registered exploratory analyses using our preferred model specifications to investigate interaction effects based on students' grade level, prior knowledge, and sex (Table A6). Separate analyses were then conducted using our preferred model specifications to estimate treatment effects for the subgroups of students: 4th graders, 5th graders, students with below average prior knowledge, students with above average prior knowledge, males, and females (Figure A2). A median split was used on the overall math composite at pretest (average standardized score) to divide students into above and below average prior knowledge groups.

**Table 4**
Estimated Treatment Effects.

| | Sample with Post-test: No Covariates ($N$ = 334) | | | Pretest Average Composite & Grade Covariates ($N$ = 316) | | | Not Missing 90 % of Post Untimed Test & Covariates ($N$ = 314) | | |
|---|---|---|---|---|---|---|---|---|---|
| Standardized Post-test Outcome | *b* | *(SE)* | *p* | *b* | *(SE)* | *p* | *b* | *(SE)* | *p* |
| Overall composite | 0.19 | (0.31) | 0.54 | 0.37 | (0.10) | 0<.01** | 0.37 | (0.10) | 0<.01** |
| Near transfer composite | 0.28 | (0.29) | 0.34 | 0.44 | (0.09) | 0<.001*** | 0.45 | (0.09) | 0<.001*** |
| Far transfer composite | −0.03 | (0.27) | 0.92 | 0.13 | (0.09) | 0.19 | 0.13 | (0.09) | 0.20 |
| Timed fraction to decimal conversion | 0.09 | (0.45) | 0.84 | 0.36 | (0.14) | 0.02* | 0.36 | (0.13) | 0.02* |
| Timed decimal to fraction conversion | 0.07 | (0.34) | 0.84 | 0.28 | (0.10) | 0.02* | 0.28 | (0.10) | 0.02* |
| Timed fraction addition | −0.03 | (0.23) | 0.89 | 0.05 | (0.15) | 0.74 | 0.04 | (0.15) | 0.79 |
| Untimed fraction to decimal conversion | 0.58 | (0.39) | 0.16 | 0.80 | (0.19) | 0<.001*** | 0.81 | (0.18) | 0<.001*** |
| Untimed decimal to fraction conversion | 0.13 | (0.27) | 0.64 | 0.28 | (0.12) | 0.03* | 0.29 | (0.12) | 0.03* |
| Untimed fraction and decimal addition | 0.27 | (0.35) | 0.45 | 0.48 | (0.17) | 0.01* | 0.49 | (0.17) | 0.01** |
| Reversed natural log of PAE 0 to 1 | 0.32 | (0.26) | 0.23 | 0.43 | (0.15) | 0.01* | 0.43 | (0.16) | 0.01* |
| Reversed natural log of PAE 0 to 5 | 0.11 | (0.33) | 0.75 | 0.26 | (0.11) | 0.03* | 0.26 | (0.10) | 0.02* |

*Note.* Percent absolute errors (PAE) on the number lines were transformed using natural log and reverse coded so that positive scores indicate better performance. Standardized scores are reported to allow for comparison across measures. The post-test standardized scores were calculated using the average grade standard deviation for the control group. The covariate is the average z-score of child performance on the pretests using the average grade standard deviation for all participants. Clustered standard errors by teachers are in parentheses. The second model is our pre-registered model (https://osf.io/kjqmz) with pretest average composite & grade covariate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

## 7. Results

### 7.1. Co-design vignettes

To answer our first research question about the key adaptations that altered the Fraction Bal intervention through co-design with teachers, we present vignettes highlighting three key modifications we made to the intervention. These adaptations include: (1) sequencing lessons to orient students, (2) roles to accommodate large class sizes, and (3) classroom lessons that connect to games on the court.

During co-design sessions, teachers emphasized the need to orient students to the games, the mathematical representations on the court, and expectations about roles and materials in the classroom rather than on the outdoor court. One teacher explained,

I'm thinking, actually, that explaining it on the board or under my document camera first and going over, like, the game, before we even go down onto the court. 'Here's how it's going to work. There's a number line…' Because they don't even really know they see it there. But they don't really know what's going on with it yet you know.

Teachers explained that students were excited when they got on the court, making it challenging to introduce the games and provide directions. Frontloading instructions in the classroom could help scaffold the students' experience and improve classroom management. One example of this was teachers' adaptations to the first two activities (in the classroom and on the court) to exclusively focus on orienting students to the design of the court and how the games worked. In the first classroom lesson, teachers decided to introduce basketball, discuss the differences between a regular court and a Fraction Ball court, and ask students to construct the court. In the second lesson, on the court, teachers introduced the game Simon Says (e.g., Simon Says: Go to 1/4, Go to 1/4 + 2/4 on the number line, Hop to the smallest decimal on the court) in the classroom before going outside. Once the class understood the rules, teachers would have the class go outside, explore the court, play Simon Says to get familiar with the math on the court, and spend the last few minutes practicing dribbling and shooting. Teachers replaced an existing game with Simon Says and created a complimentary classroom lesson in an effort to create a smoother transition from the classroom to the court, where teachers expected student management would be more challenging. Teachers decided that all games should be introduced inside before going outside. They also created a variety of materials to prepare students for the transition to the court and the new games they were learning, including laminated cards describing the roles of the game and how students move through them and a basket that holds all the items needed for the game (e.g., jerseys, hula hoops, basketballs, clipboards, etc.) to minimize transition time.

A second adaptation to the games on the court came from the challenge of accommodating larger class sizes of up to 30 students. To address this challenge, we added additional roles to the game–including a cheering section where students were responsible for supporting their classmates while they shoot–and asked students to double up in certain roles. In our co-design sessions, teachers proposed creating a new role, the "tracker," who would hold a clipboard with a court diagram and keep track of all the shots their classmates made. One teacher explained that her class was used to having roles in the classroom and that tracking was something students were already familiar with,

So they're used to roles in their small groups. That way, when they're assigned a job, they know what they're supposed to do with that job. I think that does help … So, like in math class, when there's a reporter or recorder, the usual jobs, it's like that kid knows they have to write everything down.

The tracker role had several logistical and conceptual benefits. First, it accomplished the primary goal of accommodating more students on the court using a role that many teachers already use in their classrooms. Second, the resulting data provided opportunities for students to reflect on their performance on the court and make strategy decisions for future activities (e.g., we made most of our points from the $^1/_3$ line, so we should shoot more of those in the next game). Lastly, the "tracked" points provided an opportunity for teachers to connect concrete representations of rational number values students experience on the court and hone in on procedural aspects of arithmetic and estimation in-classroom activities where students tallied the points from their games.

The most significant adaptation to the Fraction Ball intervention was the addition of classroom lessons to connect and complement the games. The previous iteration of Fraction Ball consisted of six outdoor activities (see supplemental materials in Bustamante et al., 2022 for a full description). Through our co-design with teachers, six classroom activities were added. This adaptation emerged from the feedback from teachers who delivered the pilot intervention as well as our co-design teachers. One teacher shared, "It's a waste to cover Fraction Ball without direct connections to [what they are doing in] math class." Teachers shared that classroom lessons could enhance and practice the math that students were learning in the games on the court but they needed to directly connect to what students were doing in the games. Two of the classroom lessons included a new set of activities that teachers created where students watched video highlights of WNBA/NBA games to estimate the value of shots (e.g., what would that shot be worth on a thirds or fourths court?) and track the total points made. Drawing on students' interest in local professional basketball teams and watching sports during math class, these activities taught students how to track points in the classroom prior to going to the court. Thus, tracking became the focus of several classroom activities where students were guided to make connections between estimating made shots on a Fraction Ball court printed on paper, tallying their points in a table, and counting the total scores on a number line. These three representations (i.e., diagram, table, number line) presented across multiple lessons afforded students opportunities to make deeper connections between fraction and decimal values, reinforce magnitude estimation skills, and afforded teachers to guide their meaning-making by superimposing an arithmetic sentence on a number line.

We highlight these examples across outdoor court and indoor classroom contexts to demonstrate how teachers' knowledge of their students, classroom, and schoolyard impacted the learning and engagement opportunities across the intervention. Teachers advocated for a slower pace of delivery, as reflected by the first adaptation. They requested more time to explore concepts by breaking down activities into multiple days or reducing the number of concepts introduced in each lesson/activity. These examples also highlight that teachers prioritized knowledge transfer between the court and the classroom and emphasized the importance of explicitly making connections between mathematical procedures and conceptual understanding in their lessons. We believe these co-designed teacher contributions and adaptations were largely responsible for the increased efficacy and usability of the Fraction Ball intervention observed in this study.

## 8. RCT evaluation on the impacts of fraction Ball on Students' rational number outcomes

To answer our second research question, whether the Fraction Ball intervention improved students' rational number outcomes, we present our descriptives and regression analyses on the standardized composite scores and subtests.

Descriptive statistics are presented in Table 1 for students' raw scores by subtest and in Table A2 for standardized composite scores. Separate regression analyses on raw scores for subtests and on standardized composite scores, clustering errors by teacher, showed the intervention and control groups were not statistically different in math performance at pretest, attrition, pretest missingness, grade, ethnicity, ELL status, free or reduced-price lunch, and student disability (all; Table 1 and Table A2). However, based on the district-provided information available on students' sex assigned at birth (referred to as sex, hereafter), the intervention group was statistically more likely to include female students than the control group (b = 0.13, $p = 0.002$). Also, fifth graders ($M = 0.12$, $SD = 0.61$) scored higher on the overall standardized composite score than fourth graders ($M = -0.07$, $SD = 0.83$, $t = -2.49$, $p = 0.01$). We also observed sex differences at the pretest, such that males ($M = 0.13$, $SD = 0.79$) had higher scores than females ($M = -0.10$, $SD = 0.61$, $t = -2.86$, $p = 0.005$).

Table 4 shows the impact estimates of the Fraction Ball intervention from our preferred specification shown in the second column, regression models clustering standard errors by teacher. Posttest intervention effects favored the treatment group for nearly every subtest but were not statistically significant without covariates ($.16 \leq p \leq 0.92$). Large standard errors (0.31 for overall composite) suggest realistically sized impacts could not be detected. Adding math performance at pretest and grade level as covariates, reduced standard errors (0.10 for overall composite), and the effect of the intervention became larger and statistically significant on all but two subtests. For standardized composite scores, there was an overall statistically significant impact of the intervention on the overall posttest score (b = 0.37, $p = 0.002$) and near transfer (b = 0.44, $p < 0.001$), but the impact for far transfer was not statistically significant (b = 0.13, $p = 0.19$). These findings are similar to a previous implementation of Fraction Ball (Bustamante et al., 2022), suggesting the intervention had an impact on rational number items that were embedded in the intervention and not on general to all rational number knowledge.

Fig. 2 shows the treatment estimates of the Fraction Ball intervention on all the standardized outcomes. Across standardized subtests, impact estimates were between 0.26 SD and 0.80 SD and statistically significant ($.001 < p \leq 0.03$), except for the timed fraction addition subtest ($b = 0.05$, $p = 0.74$; Table 4). Specifically, impacts on the timed fraction to decimal and decimal to fraction conversions were 0.36 SD ($p = 0.02$) and 0.28 SD ($p = 0.02$), respectively, showing statistically significant small impacts relative to other outcome measures.

A greater range of effect sizes were found on the untimed subtests. Statistically significant effects were seen on the decimal to fraction conversion subtest (b = 0.28, $p = 0.03$) and on the 0 to 5 fraction number line estimation subtest (b = 0.26, $p = 0.03$). Meanwhile, larger statistically significant impacts were found on the fraction and decimal addition subtest (b = 0.48, $p = 0.01$) and on the 0 to 1 fraction number line estimation subtest (b = 0.43, $p = 0.01$). Finally, a statistically significant effect was present on the untimed fraction to decimal conversion subtest (b = 0.80, $p < 0.001$), which was the largest impact across subtests and composite scores. Impact estimates were almost identical when excluding students who had missing data for 90 % or more of the untimed test items (see Table 4). Furthermore, Figure A1 shows students' performance at pretest and posttest by teacher and treatment group, showing variability of average student performance across teachers/classrooms and generally steeper slopes in performance growth for those who participated in Fraction Ball.

### 8.1. Robustness checks

To probe the robustness of our estimates, we also estimated GLM models with random intercepts by teacher and regressions that used FIML to account for missing data. Table A3 shows estimated treatment effects from the random intercept models, which were very similar in magnitude to the estimates from the regression models clustering standard errors at the teacher level and followed the same pattern of statistical significance/non-significance. Table A4 displays estimated fixed and random effects components for the random intercept model with the overall composite score as the outcome. Results from the baseline model indicated statistically significant variability on students' overall rational math skills and larger than expected at the teacher level (ICC = 0.56, $p < 0.01$) and student level within classrooms (ICC = 0.64, $p < 0.64$). Fortunately, because the pretests were so highly associated with the posttests, standard errors dropped dramatically when they were entered into the model. Furthermore, 43 % of the residual variance in students' scores was due to differences between teachers/classrooms. Once covariates were added to the model, the amount of residual variance due to teacher differences was reduced to 16 %, and variability between teachers/classrooms and students within-classroom remained statistically significant but were smaller in magnitude. Accounting for the random effects and controlling for students' pretest composite score and grade level, the impact estimate of Fraction Ball on the overall standardized composite score was 0.36 SD ($p < 0.001$), nearly identical to the estimate in the model that used clustered standard errors instead of random intercepts to account for the nonindependence of students within teachers.

Results from the models using FIML to get estimated effects on the full sample ($N = 360$), including students with missing data, also produced estimates following the same pattern in magnitude and significance as the previous models. Specifically, we found statistically significant effects on the overall composite and near transfer composite but not on the far transfer composite and on all subtests except timed fraction addition (see Table A5).

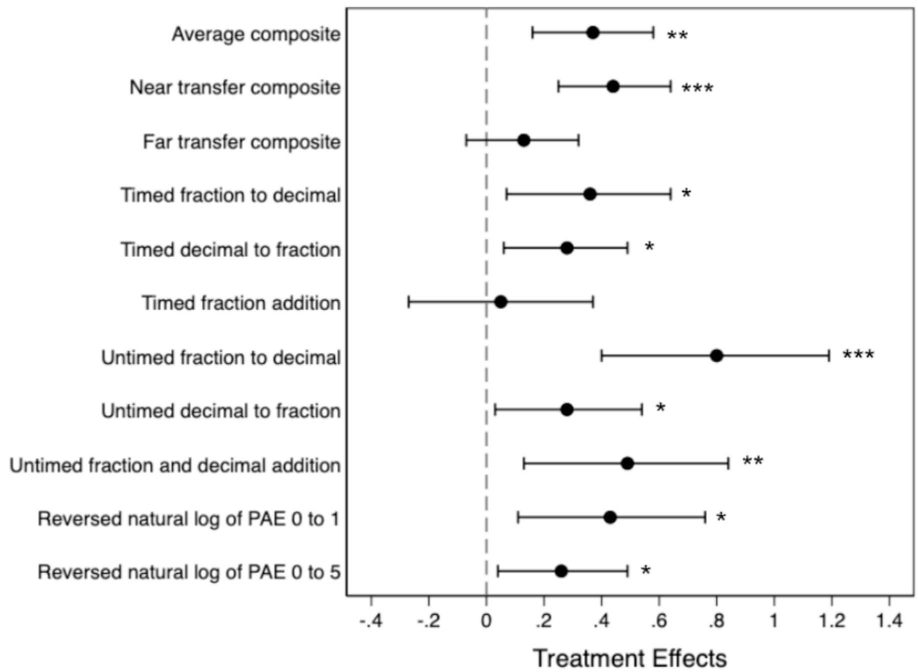



**Fig. 2.** *Current Study Posttest Treatment Estimates by Standardized outcome. Note.* Estimates represent regression coefficients for treatment status. The estimates can be interpreted as standardized treatment effects in pooled by grade level SDs of the control group. Models are specified using clustered standard errors by teachers, controlling for pre-test average composite and grade level ($N = 316$). Bars show 95 % confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

### *8.2. Moderation of treatment impacts*

Finally, we conducted pre-registered exploratory regression analyses clustering standard errors by teacher to investigate whether child characteristics–grade level, prior knowledge, and sex–moderated the effects of the intervention. First, we conducted separate regression models adding the interaction term between treatment and each of the potential moderators. As seen in Table A6, we did not find statistically significant moderation effects on the overall standardized composite score between treatment and grade (b = 0.08, $p = 0.68$) or prior knowledge (b = 0.28, $p = 0.08$). However, sex moderated the effects of the Fraction Ball intervention on the overall standardized composite scores, such that the intervention was statistically more beneficial for males than females (b = 0.21, $p = 0.03$). Next, we conducted separate regression analyses by subgroup. Figure A2 shows estimated treatment effects for 4th graders (b = 0.35, $p = 0.05$), 5th graders (b = 0.47, $p = 0.001$), students whose prior knowledge at pretest was above (b = 0.53, $p = 0.001$) and below average (b = 0.25, $p = 0.049$), males (b = 0.49, $p < 0.001$), and females (b = 0.29, $p = 0.009$).

Taken together, all pre-registered analyses with different assumptions converge to indicate Fraction Ball had a statistically significant impact on students' overall rational skills in the range of what we predicted, an effect that seems to be driven by the impact on near transfer items. Additionally, Fraction Ball had statistically significant impacts on different rational number skills. Nonetheless, we identified transfer and timed fraction addition items as areas for improvement in future iterations of the intervention. Furthermore, we will investigate whether differential treatment effects on subgroups might have been due to engagement and will make changes to the intervention accordingly to attempt to have equal impacts across all subgroups.

## 9. Discussion

Results from our RCT evaluation reflect that students in the Fraction Ball program developed a more sophisticated mental representation of rational numbers compared to students who received business-as-usual math and PE instruction. Fraction Ball students outperformed the control group in 10 of 12 rational number outcomes. Our main outcomes and subtests (i.e., number line estimation, conversions, and addition) included measures that were a combination of near and far transfer items. Although impacts on far transfer were in the positive direction, the impact was $^1/_3$ the size of the overall impact and did not reach statistical significance, indicating a need to attend in future work to students' knowledge of fraction and decimal representations not shown in Fraction Ball. This future work addressing a broader array of rational number knowledge will be crucial for wider adoption of Fraction Ball as educators will want such a substantial time investment to result in improvements on broader classroom and standardized assessments.

Although we were underpowered to detect treatment effects by subgroup (e.g., prior knowledge, sex, and grade), our exploratory analyses indicate that males improved more than females. Positive impacts on our main outcome were observed for all student groups (prior knowledge, sex, and 5th grade), except for 4th graders (although the non-significance of the treatment effect for this subgroup [p = 0.051] appears to be mostly attributable to the lower precision of the estimate rather than due to having a smaller effect size than our other outcome measures; see Fig. 2A for 4th grade estimates).

### *9.1. Expansion of fraction Ball through the DBIR/RCT approach*

Through our DBIR approach to intervention refinement and design, we made substantial improvements to the program, which strengthened its impacts even when implemented at a larger scale, compared to the first iteration administered at a single site where greater control is possible (Bustamante et al., 2022). Specifically, compared to our previous smaller-scale evaluations using within classroom randomization, we find that impacts in the current study were larger for number line estimation subtests and fraction and decimal addition (See Figs. 1 and 2; Also, see Bustamante et al., 2022). Impacts for our other subtests are comparable to those of our pilot, although our error terms are smaller. Although we cannot attribute these changes to specific modifications in the intervention, we attribute these improvements to the methodological approach of centering educators' voices in the process of co-designing our intervention, resulting in significant changes in the content, delivery, design, and dosage of Fraction Ball.

Another study by Yeager et al., (2016) employed a method akin to the two-step approach of DBIR and RCT. In their study, researchers met with their users both one-on-one and in groups (comprising 2–10 9th graders) and then piloted their intervention revisions with a small group of 20–25 9th graders. The objective was to refine an existing two-session self-administered intervention aimed at enhancing growth mindset. Yeager et al., (2016) examined five distinct features derived from these meetings in a multi-arm RCT with 3,004 MTurk participants. Subsequently, they included factors that demonstrated positive impacts in their revised intervention. In a follow up study administered at school sites, with a census of 7,501 9th graders, they found that the revised intervention outperformed the original intervention across various growth mindset measures.

It is laudable to engage in rapid causal testing with large samples through multi-arm RCTs, utilizing MTurk to evaluate which design factors lead to improvements. We used a similar approach to develop a school-based intervention. Fraction Ball originated by centering a local school's needs–to paint a basketball court and tackle persistent student challenges with rational number concepts (Bustamante et al., 2022).

The challenge arises in reconciling community needs with the methodological rigor of an RCT. For instance, the largest adaptation came from the expansion of the Fraction Ball program from exclusively outdoor activities to adding six in-classroom activities. The extension to the classroom afforded the co-design team opportunities to address major concerns educators had about the program. For example, preparing students to play the Fraction Ball games before they go outside and integrating concrete, rational number concepts presented on the court with more abstract procedures presented in the classroom. The role of "tracker" was a teacher adaptation that supported this connection between the court and the classroom by encouraging children to collect data during game play and allowing for the games on the court to accommodate more students. Yet, when we implement RCTs to evaluate the impact of our design adaptations, we allow for teacher flexibility and autonomy that RCTs traditionally discourage and frame as threats to internal validity. For example, teachers may opt to break a lesson into two separate lessons or skip certain components of activities if they are not a strong fit with their students. Instead of discouraging teacher adaptations or framing them as poor implementation fidelity, we document them and interview teachers to understand their rationale. These qualitative insights undergo thematic analyses and we apply our learnings into future iterations of the curriculum. Resulting qualitative evidence is reported to provide readers with rich descriptions of teachers' and students' experiences, context, and needs, which provide opportunities for making conjectures and theorizing about learning principles and the factors involved in scaling (Lawrence et al., 2023).

### 9.2. Contributions of results to the broader literature

A plausible mechanism for students' improvements in magnitude understanding comes from embedding concrete representations of the number line on the court and in the classroom. As the integrated theory of the number line predicts (Siegler et al., 2011), students may have engaged in making connections between non-symbolic and symbolic knowledge and, in turn, developed more accurate mental representations of whole numbers and fraction/decimal magnitudes along a mental number line. These possible cognitive changes are reflected in our data showing Fraction Ball students' performance increased in 0 to 1 and 0 to 5 number line estimation. Processes of association and analogical reasoning are deeply involved in building mental schemas that integrate whole numbers and fractions (Braithwaite & Siegler, 2021).

Analogical reasoning theories posit that reasoners build mental conceptual schemas by mapping and aligning from a familiar representation to a less familiar representation by comparing and contrasting cases (Begolli & Richland, 2016; Gentner, 1983; Gick & Holyoak, 1983; Star et al., 2016). In Fraction Ball, students are expected to map the physical magnitudes visible on the court and the life-size number line to the fraction and decimal symbols that represent them. Additionally, the fraction and decimal notations were placed side-by-side, affording students opportunities to map between the equivalent magnitude representations in the two notation formats. These theories are corroborated with evidence that Fraction Ball students improved their ability to add fractions with decimals and convert between fractions and decimals.

The combination of concrete and abstract representations enabled by in-class activities and the "tracker" role, align with several areas of psychological research on learning. There is ample evidence indicating that including both concrete and abstract representations leads to larger learning gains in mathematics compared to interacting with only abstract or only concrete representations (Fyfe et al., 2012; Goldstone & Son, 2005; Kokkonen & Schalk, 2021). However, making connections between two or more representations is challenging. This complexity likely arises from the potential demand on executive function resources required to maintain these representations in mind and the prior knowledge required to identify alignments and misalignments (Begolli et al., 2018). Thus, instructional supports that make explicit connections between representations are often needed for most reasoners to successfully build mental schemas and avoid potential misconceptions (Begolli & Richland, 2016; Richland et al., 2007).

It may be the case that rich instructional supports embedded in Fraction Ball leveled the playing field between students with lower and higher prior knowledge, suggesting our intervention accounted for varied student abilities. In particular, we posit that in-class activities and the "tracker" sheet and role served to elucidate the connections between abstract and concrete representations and in turn conceptual and procedural knowledge, in several ways. For example, the tracking sheet supported students to draw a connection between the concrete Fraction Ball design on the court and its abstract representation on the tracking sheet. In turn, this may have facilitated students drawing connections between multiple abstract and concrete representations of fractions and decimals, such as requiring them to translate fraction magnitudes on the court, to number lines (on court and in-class), summary tables (in-class), and addition procedures (in-class). Moreover, to reinforce these connections, activities were sequenced to cycle between the court (rich in concrete representations) and the classroom (rich in abstract representations), in turn reinforcing students' conceptual and procedural knowledge in a cyclical fashion (Rittle-Johnson et al., 2001).

Overall, Fraction Ball draws from multiple learning principles, and with our data and design, it was not feasible to isolate specific components that explained student learning. Future studies should disentangle possible mechanisms that impact learning for program improvement.

Understanding how interventions affect student outcomes more broadly and across studies has been challenging with diverse researcher-developed measures (as exemplified by Kraft, 2020) and distinct student populations. To provide context for the magnitude of our impacts, we juxtapose our effect sizes with relevant preceding interventions. For instance, Fazio et al.'s (2016) intervention involving fraction number lines was less intensive, encompassing a 15-minute training session and they found an improvement ranging from 0.13 SDs to 0.58 SDs. This range is similar to some of our findings, although the tasks in our current study are more distanced from the intervention in terms of their format, context, and timing.

On the other hand, in a 12-week (36-unit) intensive intervention carried out by grant contracted tutors, Fuchs et al. (2016) observed impacts of around one standard deviation in number line estimation, between one and 2.5 standard deviations in fraction calculation, and between 0.4 and 0.9 standard deviations on a selection of pertinent problems derived from the National Assessment of Educational Progress. These impacts likely reflect the considerable intensity and efficacy of these programs (which consisted of direct instruction, high dosage, and grant funded non-school staff to deliver the intervention), along with the alignment of the program's content with some of the outcomes.

### 9.3. Limitations

From a mechanistic approach, the primary limitation of this work stems from what is also a strength of the work – an intervention with deeply integrated components. This meant that the relationship between specific intervention components added to the current iteration could not be directly linked to the stronger measured effects, though we acknowledge this would be useful for broader RCT interventions focused on disentangling the causal factors underpinning the learning gains.

In addition, we did not have the opportunity to explore the sociocultural context for the results and context variations seem to differ in how they prompt students' motivation (Yeon Lee et al., 2024). Importantly, both sexes improved from our intervention, but male students improved more than their female counterparts. One possibility is that sociocultural factors including, perhaps, popular perceptions of basketball as a male dominant sport may have waned the engagement levels and in turn softened the impacts of the intervention for females. To promote higher engagement for female students, in the current iteration of Fraction Ball, students were introduced to the Women's National Basketball League (WNBA) and watched short clips of the WNBA and estimated the value of shots pretending they were playing on a Fraction Ball court. Although we do not have data to investigate all sociocultural factors, in related work, we found no sex differences on the impacts of Fraction Ball on students' emotions toward math (Guo et al., 2024). Further, in our pilot study, we did not find that sex moderated the impacts of the intervention (Bustamante et al., 2022). Given our lack of a priori hypothesis about moderation by sex and these previous findings, it is possible this interaction was a false positive. Still, future iterations of the program should proactively promote intervention features that support engagement for female students.

## 10. Conclusion

This study serves as a model for how to develop interventions that capitalize on accumulated knowledge from psychological science and integrate it with educators' on-the-ground expertise and experience. Whereas this is typical in DBIR, few interventions created using co-design are subsequently evaluated using rigorous methods like RCTs, questioning the causal links and limiting large-scale dissemination and the generalizability of the theoretical insights derived. In contrast, our RCT evaluation has the potential to provide important insights into the interplay of cognition within authentic school settings influenced by dynamic social, cultural, and structural factors. This study offers a rigorous methodological approach for leveraging insights from psychological science and integrating with pedagogical practices through co-design to align theory and practice within impactful interventions in real-world contexts.

## CRediT authorship contribution statement

**Kreshnik N. Begolli:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Vanessa N. Bermudez:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Formal analysis, Data curation. **LuEttaMae Lawrence:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Formal analysis. **Lourdes M. Acevedo-Farag:** Writing – review & editing. **Sabrina V. Valdez:** Project administration. **Evelyn Santana:** Project administration. **Daniela Alvarez-Vargas:** Supervision, Project administration, Conceptualization. **June Ahn:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Drew Bailey:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Katherine Rhodes:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Lindsey E. Richland:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Andres S. Bustamante:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We would like to thank the teachers, students, and administrators across six schools in the Santa Ana Unified School District for their partnership, creativity, and enthusiasm. They made this work possible. This work was supported by the Advanced Education Research Development Fund (AERDF) through their EF + Math Program [grant number FRACT-0620].

## Appendix A

**Table A1**
Descriptive Statistics for Demographic Variables and Raw Scores of Outcomes by Grade.

| | Grade 4 [a, b] | | | | | | Grade 5[c, d] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Full | | Control | | Intervention | | Full | | Control | | Intervention | |
| Variable | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD |
| ***Pre-tests*** | | | | | | | | | | | | |
| Timed fraction to decimal conversion | 6 % | 16 % | 3 % | 7 % | 9 % | 19 % | 15 % | 28 % | 29 % | 35 % | 3 % | 10 % |
| Timed decimal to fraction conversion | 20 % | 23 % | 17 % | 20 % | 22 % | 25 % | 19 % | 25 % | 24 % | 25 % | 13 % | 24 % |
| Timed fraction addition | 45 % | 26 % | 41 % | 24 % | 49 % | 28 % | 51 % | 32 % | 58 % | 32 % | 45 % | 31 % |
| Untimed fraction to decimal conversion | 5 % | 14 % | 2 % | 7 % | 7 % | 17 % | 14 % | 27 % | 27 % | 32 % | 2 % | 10 % |
| Untimed decimal to fraction conversion | 44 % | 47 % | 35 % | 44 % | 50 % | 48 % | 31 % | 40 % | 38 % | 40 % | 25 % | 40 % |
| Untimed fraction and decimal addition | 19 % | 32 % | 15 % | 27 % | 23 % | 35 % | 33 % | 43 % | 58 % | 47 % | 11 % | 22 % |
| PAE fraction number line 0 to 1 | 23 % | 18 % | 27 % | 18 % | 21 % | 19 % | 19 % | 16 % | 16 % | 17 % | 21 % | 15 % |
| PAE fraction number line 0 to 5 | 29 % | 15 % | 30 % | 14 % | 28 % | 16 % | 26 % | 16 % | 20 % | 15 % | 31 % | 15 % |

(continued on next page)

**Table A1** (*continued*)

| | Grade 4 [a, b] | | | | | | Grade 5 [c, d] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Full | | Control | | Intervention | | Full | | Control | | Intervention | |
| Variable | M | SD | M | SD | M | SD | M | SD | M | SD | M | SD |
| ***Post-tests*** | | | | | | | | | | | | |
| Timed fraction to decimal conversion | 16 % | 29 % | 5 % | 12 % | 24 % | 34 % | 29 % | 36 % | 40 % | 41 % | 19 % | 28 % |
| Timed decimal to fraction conversion | 26 % | 28 % | 16 % | 22 % | 33 % | 30 % | 32 % | 32 % | 42 % | 38 % | 24 % | 24 % |
| Timed fraction addition | 56 % | 23 % | 51 % | 23 % | 59 % | 22 % | 56 % | 33 % | 64 % | 35 % | 49 % | 30 % |
| Untimed fraction to decimal conversion | 27 % | 39 % | 8 % | 22 % | 42 % | 43 % | 42 % | 40 % | 43 % | 43 % | 41 % | 37 % |
| Untimed decimal to fraction conversion | 46 % | 46 % | 33 % | 46 % | 55 % | 45 % | 50 % | 46 % | 60 % | 46 % | 42 % | 45 % |
| Untimed fraction and decimal addition | 33 % | 39 % | 16 % | 29 % | 45 % | 42 % | 49 % | 42 % | 58 % | 44 % | 41 % | 39 % |
| PAE fraction number line 0 to 1 | 20 % | 18 % | 25 % | 18 % | 16 % | 17 % | 14 % | 14 % | 14 % | 14 % | 14 % | 15 % |
| PAE fraction number line 0 to 5 | 24 % | 15 % | 29 % | 15 % | 20 % | 14 % | 21 % | 15 % | 16 % | 13 % | 25 % | 15 % |

*Note.* PAE = percent absolute error.

[a] Grade 4 Pre-test: Full $N = 208$, Control $N = 87$, Intervention $N = 121$.

[b] Grade 4 Post-test: Full $N = 202$, Control $N = 86$, Intervention $N = 116$.

[c] Grade 5 Pre-test: Full $N = 134$, Control $N = 64$, Intervention $N = 70$.

[d] Grade 5 Post-test: Full $N = 132$, Control $N = 61$, Intervention $N = 71$.

**Table A2**

Summary Statistics of Fraction Ball Standardized Scores of Outcomes.

| | Full Sample | | | Control | | | Treatment | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Construct | *N* | Mean | SD | *N* | Mean | SD | *N* | Mean | SD | *b* | *p* |
| ***Pre-tests*** | | | | | | | | | | | |
| Overall composite | 342 | −0.00 | 0.71 | 151 | 0.08 | 0.78 | 191 | −0.06 | 0.64 | −0.14 | 0.60 |
| Near transfer composite | 342 | 0.00 | 0.66 | 151 | 0.07 | 0.73 | 191 | −0.06 | 0.60 | −0.13 | 0.62 |
| Far transfer composite | 342 | 0.00 | 0.67 | 151 | 0.08 | 0.75 | 191 | −0.06 | 0.60 | −0.14 | 0.56 |
| ***Post-tests*** | | | | | | | | | | | |
| Overall composite | 334 | 0.00 | 0.85 | 147 | −0.11 | 0.87 | 187 | 0.08 | 0.82 | 0.19 | 0.54 |
| Near transfer composite | 334 | −0.00 | 0.84 | 147 | −0.16 | 0.83 | 187 | 0.12 | 0.82 | 0.28 | 0.34 |
| Far transfer composite | 334 | −0.00 | 0.72 | 147 | 0.02 | 0.78 | 187 | −0.01 | 0.68 | −0.03 | 0.92 |

*Note.* p-value is based on a two-tailed *t*-test comparing treatment and control group on each variable, clustering standard errors by teacher. The N in the Construct section refers to the total sample possible, including student attrition. Pre-test raw scores were standardized using the average grade standard deviation from Grade 4 and Grade 5. Post-test raw scores were also standardized using the average grade standard deviation but from the control group only. Indices are only shown as standardized scores to facilitate interpretation as they contain raw scores and reverse scored natural log transformations of the percent absolute error (PAE) from the number line items.

**Table A3**

Estimated Treatment Effects from General Linear Mixed Models ($N_{teachers} = 16$).

| | Sample with Post-test: No Covariates ($N_{students} = 334$) | | | Pretest Average Composite & Grade Covariates ($N_{students} = 316$) | | | Not Missing 90 % of Post Untimed Test & Covariates ($N_{students} = 314$) | | |
|---|---|---|---|---|---|---|---|---|---|
| Standardized Post-test Outcome | *b* | *(SE)* | *p* | *b* | *(SE)* | *p* | *b* | *(SE)* | *p* |
| Overall composite | 0.25 | (0.28) | 0.36 | 0.36 | (0.10) | 0<.01** | 0.35 | (0.11) | 0<.01** |
| Near transfer composite | 0.33 | (0.26) | 0.20 | 0.43 | (0.10) | 0<.001*** | 0.43 | (0.10) | 0<.001*** |
| Far transfer composite | 0.03 | (0.24) | 0.89 | 0.12 | (0.10) | 0.24 | 0.11 | (0.10) | 0.27 |
| Timed fraction to decimal conversion | 0.20 | (0.41) | 0.63 | 0.34 | (0.15) | 0.02* | 0.34 | (0.15) | 0.02* |
| Timed decimal to fraction conversion | 0.17 | (0.31) | 0.58 | 0.29 | (0.11) | 0.01** | 0.29 | (0.11) | 0.01** |
| Timed fraction addition | −0.03 | (0.21) | 0.89 | 0.02 | (0.15) | 0.89 | 0.01 | (0.15) | 0.97 |
| Untimed fraction to decimal conversion | 0.62 | (0.36) | 0.09 | 0.76 | (0.18) | 0<.001*** | 0.76 | (0.19) | 0<.001*** |
| Untimed decimal to fraction conversion | 0.18 | (0.25) | 0.46 | 0.28 | (0.11) | 0.01* | 0.29 | (0.11) | 0.01* |
| Untimed fraction and decimal addition | 0.31 | (0.33) | 0.34 | 0.46 | (0.17) | 0.01** | 0.45 | (0.17) | 0.01** |
| Reversed natural log of PAE 0 to 1 | 0.36 | (0.25) | 0.14 | 0.43 | (0.15) | 0.01** | 0.42 | (0.16) | 0.01** |
| Reversed natural log of PAE 0 to 5 | 0.20 | (0.30) | 0.49 | 0.27 | (0.12) | 0.02* | 0.27 | (0.11) | 0.02* |

*Note.* Models are specified using random intercepts at the teacher level. Percent absolute errors (PAE) on the number lines were transformed using natural log and reverse coded so that positive scores indicate better performance. Standardized scores are reported to allow for comparison across measures. The post-test standardized scores were calculated using the average grade standard deviation for the control group. The covariate is the average z-score of child performance on the pretests using the average grade standard deviation for all participants. Adjusted standard errors are in parentheses. The second model is our pre-registered model (https://osf.io/kjqmz) with pretest average composite & grade covariates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table A4**
Generalized Linear Mixed Model with Posttest Overall Composite as the Outcome ($N_{teachers}$ = 16).

| | Baseline Model ($N_{students}$ = 334) | | | Full Sample No Covariates ($N_{students}$ = 334) | | | Pretest Average Composite & Grade Covariates ($N_{students}$ = 316) | | | Not Missing 90 % of Post Untimed & Covariates ($N_{students}$ = 314) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variable | *b* | *(SE)* | *p* | *b* | *(SE)* | *p* | *b* | *(SE)* | *p* | *b* | *(SE)* | *p* |
| ***Fixed Effects*** | | | | | | | | | | | | |
| Intercept | −0.02 | (0.14) | 0.89 | −0.15 | (0.20) | 0.46 | −0.23 | (0.08) | 0.01** | −0.22 | (0.08) | 0.01** |
| Treatment | | | | 0.25 | (0.28) | 0.37 | 0.36 | (0.10) | 0<.001*** | 0.35 | (0.11) | 0<.001*** |
| Pretest average standardized score | | | | | | | 0.96 | (0.04) | 0<.001*** | 0.95 | (0.04) | 0<.001*** |
| Grade | | | | | | | 0.07 | (0.09) | 0.46 | 0.07 | (0.10) | 0.50 |
| ***Error Standard Deviation*** | | | | | | | | | | | | |
| Teacher Intercept | 0.55 | (0.10) | 0<.01** | 0.54 | (0.10) | 0<.001*** | 0.18 | (0.04) | 0<.001*** | 0.19 | (0.04) | 0<.001*** |
| Residual | 0.64 | (0.03) | 0<.001*** | 0.64 | (0.03) | 0<.001*** | 0.42 | (0.02) | 0<.001*** | 0.42 | (0.02) | 0<.001*** |
| Intraclass correlation | 0.43 | (0.09) | | 0.42 | (0.09) | | 0.16 | (0.06) | | 0.17 | (0.07) | |

*Note.* Models are specified using random intercepts at the teacher level. Standardized scores are reported to allow for comparison across models. The average posttest composite is the mean z-score of the eight subtests using average grade standard deviations for the control group. The pretest average standardized score is the mean z-score of the eight subtests using average grade standard deviations for all participants. Adjusted standard errors are in parentheses. The third model is our pre-registered model (https://osf.io/kjqmz) with pretest average composite & grade covariates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table A5**
Estimated Treatment Effects using Full Information Maximum Likelihood for Missing Data.

| | Full Sample Pretest Average Composite& Grade Covariates | | | |
|---|---|---|---|---|
| Standardized Post-test Outcome | *b* | *(SE)* | *p* | *N* |
| Overall composite | 0.37 | (0.10) | 0<.001*** | 360 |
| Near transfer composite | 0.44 | (0.09) | 0<.001*** | 360 |
| Far transfer composite | 0.12 | (0.09) | 0.18 | 360 |
| | 0.34 | (0.13) | 0.01* | 360 |
| Timed decimal to fraction conversion | 0.26 | (0.10) | 0.01* | 360 |
| Timed fraction addition | 0.07 | (0.15) | 0.64 | 360 |
| Untimed fraction to decimal conversion | 0.78 | (0.18) | 0<.001*** | 360 |
| Untimed decimal to fraction conversion | 0.28 | (0.12) | 0.02* | 360 |
| Untimed fraction and decimal addition | 0.47 | (0.16) | 0<.01** | 360 |
| Reversed natural log of PAE 0 to 1 | 0.45 | (0.14) | 0<.01** | 360 |
| Reversed natural log of PAE 0 to 5 | 0.27 | (0.11) | 0.01* | 360 |

*Note.* Percent absolute errors (PAE) on the number lines were transformed using natural log and reverse coded so that positive scores indicate better performance. Standardized scores are reported to allow for comparison across measures. The post-test standardized scores were calculated using the average grade standard deviation for the control group. The covariate is the average z-score of child performance on the pretest using the average grade standard deviation for all participants. Clustered standard errors by teachers are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

**Table A6**
Estimated Treatment and Treatment by Moderator Effects with Controls on Overall Composite.

| Models | *b* | *(SE)* | *p* | *N* |
|---|---|---|---|---|
| ***Grade Model*** | | | | |
| Treatment | 0.34 | (0.15) | 0.04* | 316 |
| 5th grade X treatment | 0.08 | (0.19) | 0.68 | 316 |
| 5th grade | 0.04 | (0.14) | 0.75 | 316 |
| Pretest average standardized score | 1.00 | (0.06) | 0<.001*** | 316 |
| ***Prior Knowledge Model*** | | | | |
| Treatment | 0.23 | (0.12) | 0.08 | 316 |
| Above average prior knowledge X treatment | 0.28 | (0.15) | 0.08 | 316 |
| Above average prior knowledge | −0.12 | (0.13) | 0.35 | 316 |
| 5th grade | 0.12 | (0.09) | 0.21 | 316 |
| Pretest average standardized score | 0.97 | (0.09) | 0<.001*** | 316 |
| ***Sex Model*** | | | | |
| Treatment | 0.49 | (0.11) | 0<.001*** | 301 |
| Female X treatment | −0.21 | (0.09) | 0.03** | 301 |
| Female | 0.05 | (0.03) | 0.17 | 301 |
| 5th grade | 0.10 | (0.08) | 0.26 | 301 |
| Pretest average standardized score | 0.97 | (0.05) | 0<.001*** | 301 |

*Note.* The beta predictors are based on two separate models with average composite score as the outcome on treatment with controls for pretest average composite and grade plus respective interaction terms and main effects (treatment X grade; treatment X prior knowledge; treatment X sex). Standardized scores are reported to allow for comparisons across models. Clustered standard errors by teacher are in parentheses. In the Grade Model, 4th grade is the reference category. In the Prior Knowledge Model, below average prior knowledge is the reference category. The interaction remains non-statistically significant when using prior knowledge as a continuous variable. However, the effect of treatment, which was non-statistically significant becomes statistically significant when using the continuous variable ($b = 0.37$, $SE = 0.10$, $p = 0.002$). In the Sex Model, male is the reference category. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.


**Fig. A1.** Total Correct for Pre and Post-Tests by Teacher and Treatment

*Note.* Total correct is the sum of correct items across the eight subtests (44 items). For the number line estimation items, the items were scored as correct if the PAE was less or equal to 0.10, which means the error was within 10 %. The circles represent the average total correct for each teacher ($N = 16$) at pre- and post-test by treatment status. Bars show 95 % confidence intervals.

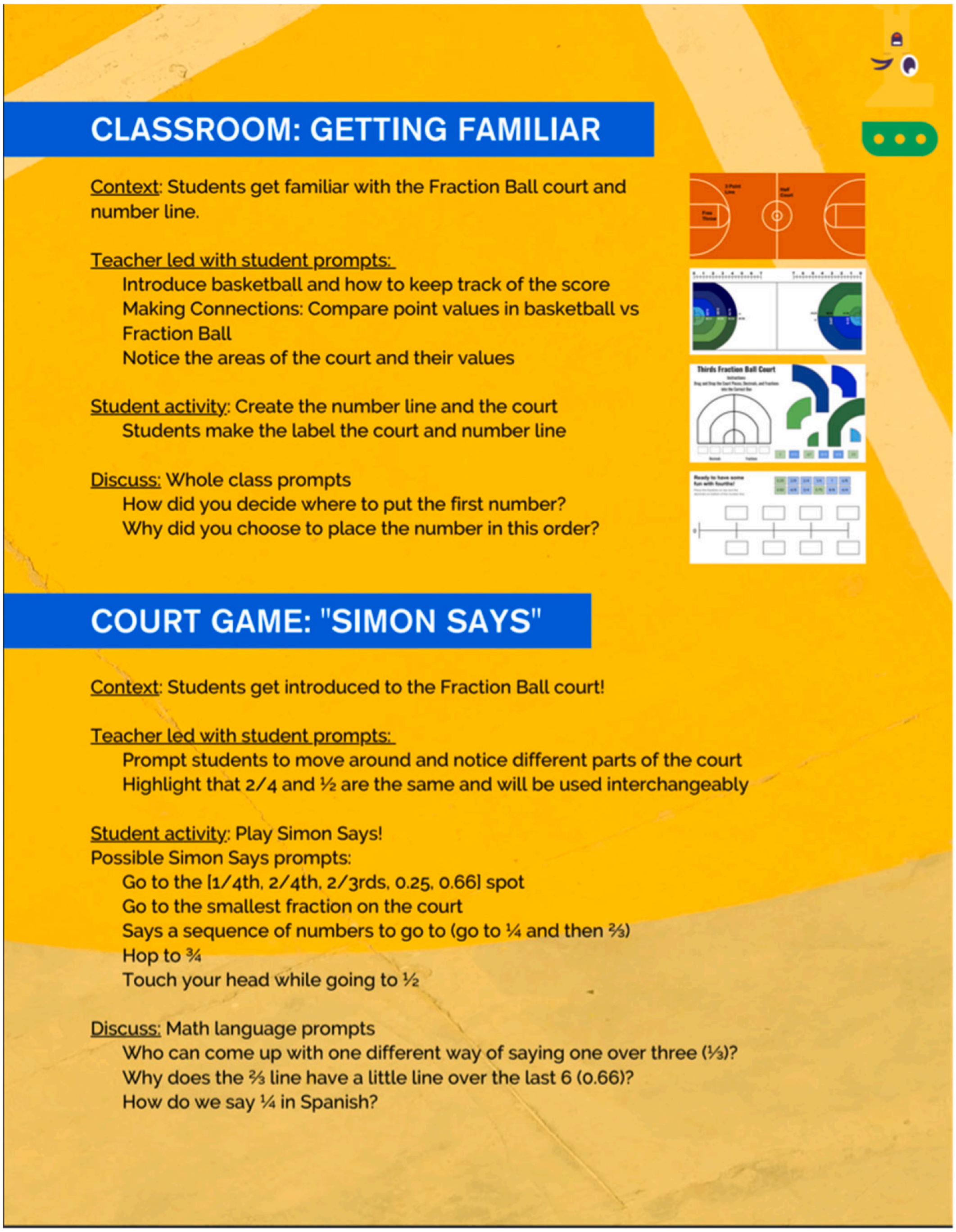



**Fig. A2.** Post-test Treatment Estimates on Overall Composite Score by Subgroup

*Note.* Estimates represent regression coefficients for treatment status on the overall composite score at post-test. The estimates can be interpreted as standardized treatment effects in pooled grade level SDs of the control group. Models are specified using clustered standard errors by teachers, controlling for pre-test average composite and grade level ($N = 316$ for grade and prior knowledge subgroups, $N = 301$ for sex subgroups). A median split on the pre-test average composite was done to create below and above average prior knowledge groups. Bars show 95 % confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

## Appendix B: Teacher activity guide

# Day 1

## BIG IDEA

Students engage in playful activities that reinforce magnitude understanding around rational numbers by familiarizing themselves with the representations on the Fraction Ball court and the number line. Day 1 activities engage students in comparing fraction and decimal notations as well as equivalent fractions.

## FRACTION BALL CONNECTIONS

Students will learn the fraction and decimal values on the Fraction Ball court and the connection between court positions, the values, and the number line. Day 1 activities will help students understand the layout of the court before moving outside. Simon Says helps students get familiar with the Fraction Ball court before going outside to play a game for the first time.

## MATERIALS

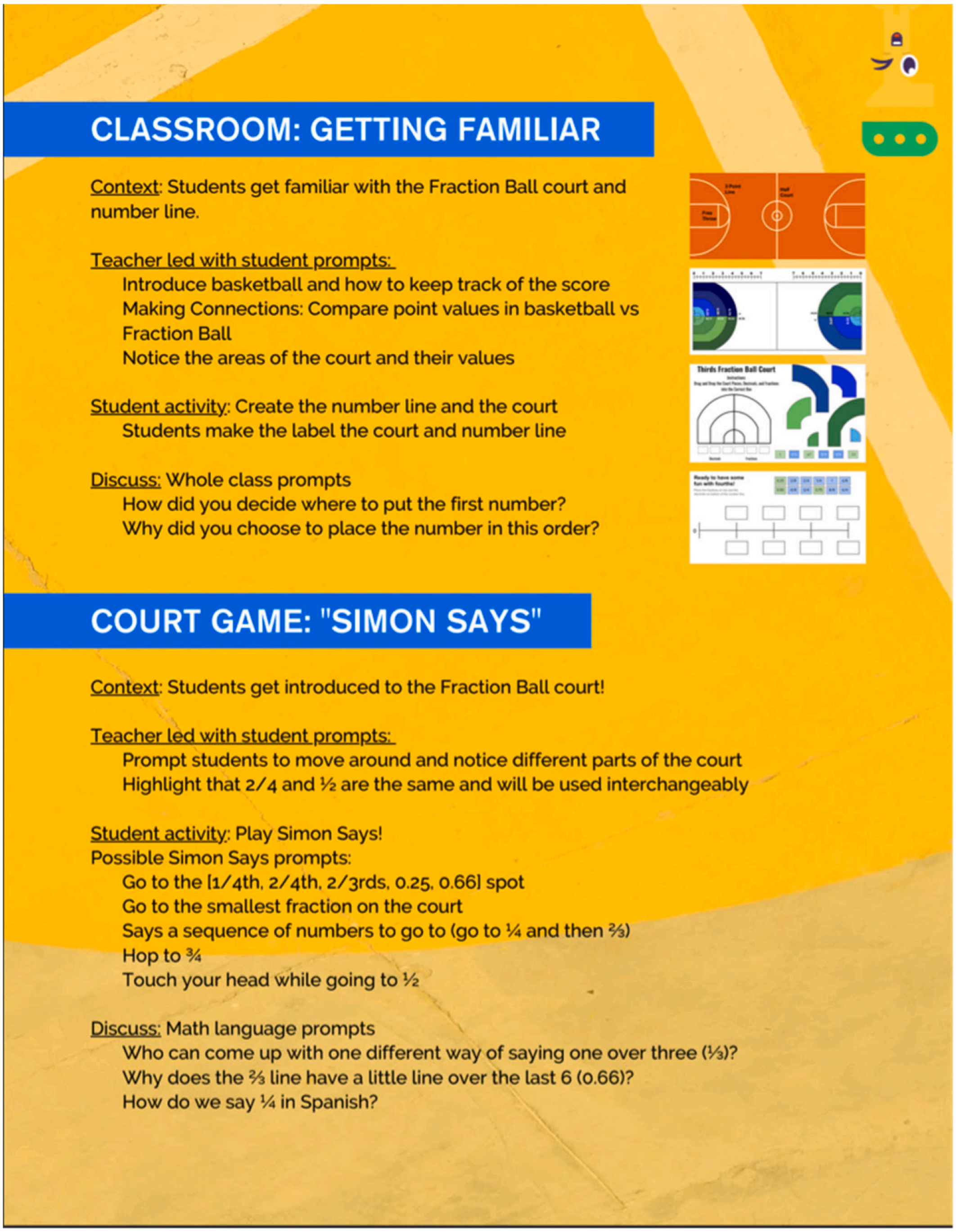
Digital Worksheet

Teaching Slides

## STANDARDS

**Math Standards: 4.NF.A.1; 4.NF.A.2; 4.NF.B.3; 4.NF.C.6; 5.NF.A.2**

.



## CLASSROOM: GETTING FAMILIAR

Context: Students get familiar with the Fraction Ball court and number line.

Teacher led with student prompts:
- Introduce basketball and how to keep track of the score
- Making Connections: Compare point values in basketball vs Fraction Ball
- Notice the areas of the court and their values

Student activity: Create the number line and the court
- Students make the label the court and number line

Discuss: Whole class prompts
- How did you decide where to put the first number?
- Why did you choose to place the number in this order?

## COURT GAME: "SIMON SAYS"

Context: Students get introduced to the Fraction Ball court!

Teacher led with student prompts:
- Prompt students to move around and notice different parts of the court
- Highlight that 2/4 and ½ are the same and will be used interchangeably

Student activity: Play Simon Says!
Possible Simon Says prompts:
- Go to the [1/4th, 2/4th, 2/3rds, 0.25, 0.66] spot
- Go to the smallest fraction on the court
- Says a sequence of numbers to go to (go to ¼ and then ⅔)
- Hop to ¾
- Touch your head while going to ½

Discuss: Math language prompts
- Who can come up with one different way of saying one over three (⅓)?
- Why does the ⅔ line have a little line over the last 6 (0.66)?
- How do we say ¼ in Spanish?

.


Day 2
BIG IDEA
Students will exercise their knowledge of the Fraction Ball court by documenting where shots are made during a Lost Angeles Sparks game and estimating their value on the Fraction Ball court. Students will reinforce their fluency in rational number addition and comparison as well as summarizing numerical sequences with tools. In addition, they engage with a life size number line on the court which can support arithmetic operations on the number line with rational numbers.
FRACTION BALL CONNECTIONS
Students will practice estimation by predicting where a shot was made from and what the fraction value might be. Students will be using their knowledge of the FB court within a new context and practice using different approaches to add shots together.
MATERIALS
Worksheet
Teaching Slides
Sparks Video
Basketballs
Pinnies
Clipboard
Tracking Worksheet
STANDARDS
Math Standards: 4.NF.A.1, 4.NF.A.2, 4.NF.B.3, 4.NF.C.6, 4.NF.C.7, 5.NF.A.1, 5.NF.A.2
PE Standards: 4.NF.3, 4.MD.2, 4.NF.5, 4.NF.7, & MP.2, MP.4, MP.5, MP.6
See script for specific targeted skills.

.


CLASSROOM: TRACK A SPARKS GAME
Context: Students become Basketball Reporters! Help the Sparks track shots with FB court, Table, and number line.
Teacher led with student prompts:
Compare points in basketball vs Fraction Ball
Extend the Number line: How many basketball courts that are 3/3's would equal 5 on the number line?
Connect tracking shots on FB court with Table tallies
Prompt students for solution strategies (skipping on number line, number sentence)
Take home message: The court, table, and number line are tools that we can use to record and calculate the total shots made!
Group Work: Worksheets for each team. ½ of teams get decimal & ½ fraction
Students work on court, table, number line-then switch roles.
Play video, pause (students mark shots) switch roles every 2 shots, repeat x 6
Make visible for students: Fraction Ball court, Table, Number line
Discuss: Take home message
FB Court shows (⅔ is twice as big as ⅓ (2 x ⅓ = ⅔)
Table = neat = less errors = flexible range (>1)
Number line is like a calculator
COURT GAME: "MAKE IT COUNT"
Each team is allowed to take exactly 15 shots from any distance they want
Each player gets 3 shots before switching roles
The goal is to get the highest amount of points possible
There is no time limit
Rule: Must take turns shooting with the other team


Day 3
BIG IDEA
Students will exercise their knowledge of the Fraction Ball court by documenting where shots are made during a Lakers game and estimating their value on the Fraction Ball court. Students will reinforce their fluency in rational number addition and comparison as well as summarizing numerical sequences with tools. In addition, they engage with a life size number line on the court which can support arithmetic operations on the number line with rational numbers.
FRACTION BALL CONNECTIONS
Students get to practice estimation by predicting where a shot was made from and what the fraction value might be. Students will be using their knowledge of the FB court within a new context and practice using different approaches to add shots together.
MATERIALS
Worksheet
Teaching Slides
Lakers Video
Basketballs
Pinnies
Clipboard
Tracking Worksheet
STANDARDS
Math Standards: 4.NF.A.1, 4.NF.A.2, 4.NF.B.3, 4.NF.C.6, 4.NF.C.7, 5.NF.A.1, 5.NF.A.2
PE Standards: 4.NF.3, 4.NF.4, MP.2, MP.4, MP.5, MP.6

.


CLASSROOM: TRACK A LAKERS GAME
Context: Students become Basketball Reporters! Help the Lakers track shots with Fraction Ball court, table, and number line.
Review or continue day 2 (if needed)
Fraction ball Court, Table, & number line
Group Work: Worksheets for each team.
½ of teams get decimal & ½ fraction
Students work on court, table, number line
Play video, pause (students mark shots) switch roles every 2 shots, repeat x 6
Make visible for students: Fraction Ball court, Table, Number line
Discuss: Take home message
The number line can tell us how big is each fraction AND Decimal just by looking at it!
The number sentence helps us keep track of many shots
COURT GAME: "EXACTLY"
Teacher calls out a number and each team has to alternate shooting baskets until one team reaches that number EXACTLY
It's not about who makes the most shots, but which team gets the EXACT number fastest
The first team to get to the exact number I call out wins
If you go past the exact number you automatically lose that round!
Rule: The shot caller must call out the shots and the counter must speak the math out loud to get the points
Add a challenge: Make the number 9/4, 5.25, 4/8, or 12/3

.

# Day 4

## BIG IDEA

Students will sum up their points made throughout the games (e.g, Make it Count, Rapid Fire) and analyze the strategies that they utilized during their game play. Students will express magnitude understanding of rational number vocabulary (e.g., In this game, students made one shot from the 3/3 line, and 2 from the ⅓ line which equals 1 ⅔) and magnitude comparison of their strategies used (e.g., In Make it Count, it was easier for our team to shoot from the 3/4ths line but in Rapid Fire, it was a better strategy to shoot closer to maximize points).

## FRACTION BALL CONNECTIONS

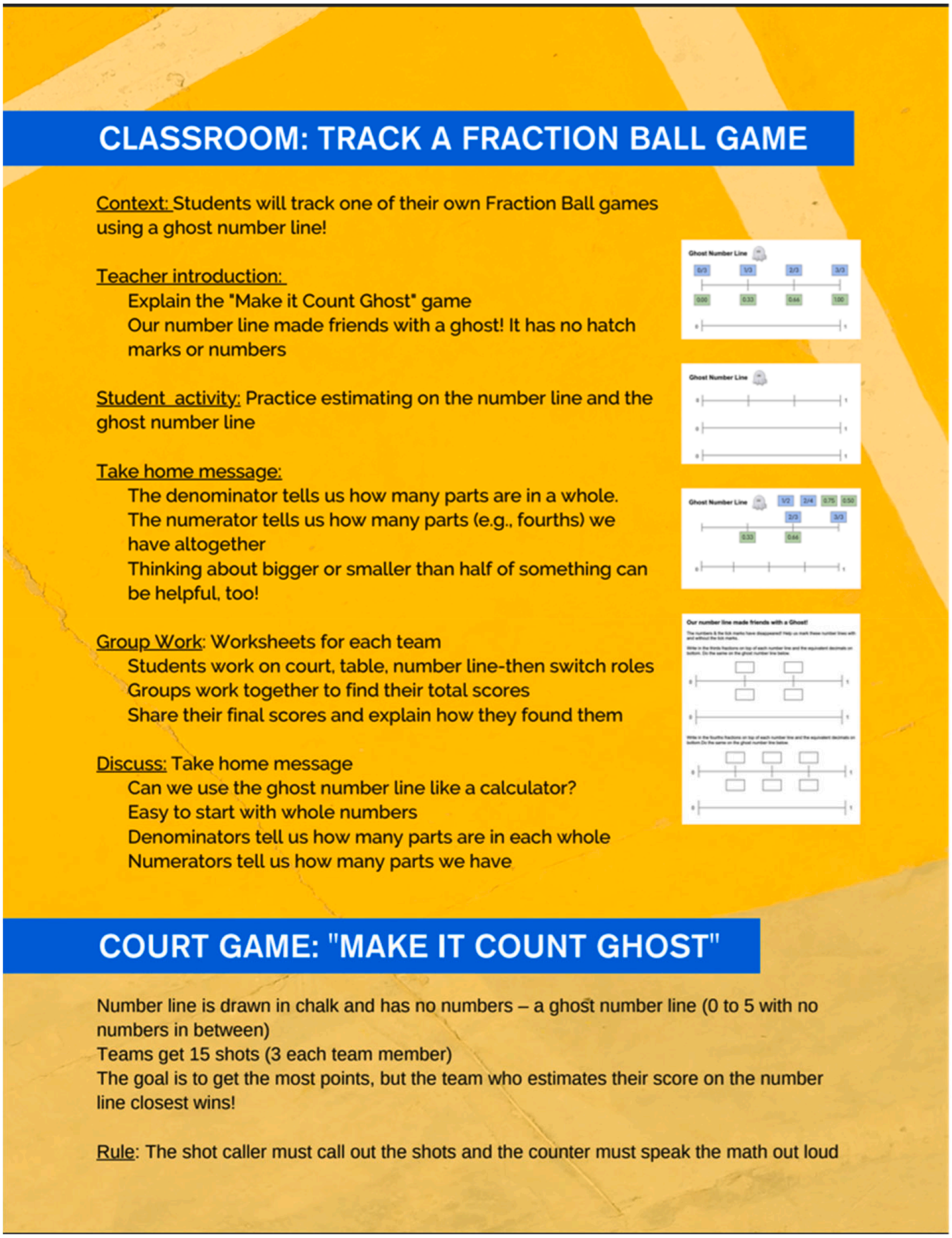
Students will use the Fraction Ball tracking worksheet and reflect on their game play and check their work.

## MATERIALS

Basketballs
Pinnies
Clipboard
Tracking Worksheet

## STANDARDS

**Math Standards: 4.NF.A.1, 4.NF.A.2, 4.NF.B.3, 4.NF.B.3, 4.NF.B.3.B, 4.NF.C.6, 4.NF.C.7, 5.NF.A.2**
**PE Standards: 4.NF.3, 4.NF.4, MP.2, MP.4, MP.5, MP.6**


CLASSROOM: TRACK A FRACTION BALL GAME
Context: Students will track one of their own Fraction Ball games using a ghost number line!
Teacher introduction:
Explain the "Make it Count Ghost" game
Our number line made friends with a ghost! It has no hatch marks or numbers
Student activity: Practice estimating on the number line and the ghost number line
Take home message:
The denominator tells us how many parts are in a whole.
The numerator tells us how many parts (e.g., fourths) we have altogether
Thinking about bigger or smaller than half of something can be helpful, too!
Group Work: Worksheets for each team
Students work on court, table, number line-then switch roles
Groups work together to find their total scores
Share their final scores and explain how they found them
Discuss: Take home message
Can we use the ghost number line like a calculator?
Easy to start with whole numbers
Denominators tell us how many parts are in each whole
Numerators tell us how many parts we have
COURT GAME: "MAKE IT COUNT GHOST"
Number line is drawn in chalk and has no numbers – a ghost number line (0 to 5 with no numbers in between)
Teams get 15 shots (3 each team member)
The goal is to get the most points, but the team who estimates their score on the number line closest wins!
Rule: The shot caller must call out the shots and the counter must speak the math out loud

.

# Day 5

## BIG IDEA

Students will sum up their points made throughout the games and analyze the strategies that they utilized during their game play. Students will express magnitude understanding of rational number vocabulary (e.g., In this game, students made one shot from the 3/3 line, and 2 from the $\frac{1}{3}$ line which equals $1\frac{2}{3}$) and magnitude comparison of their strategies used.

## FRACTION BALL CONNECTIONS

Students will use the Fraction Ball tracking worksheet and reflect on their game play and check their work.

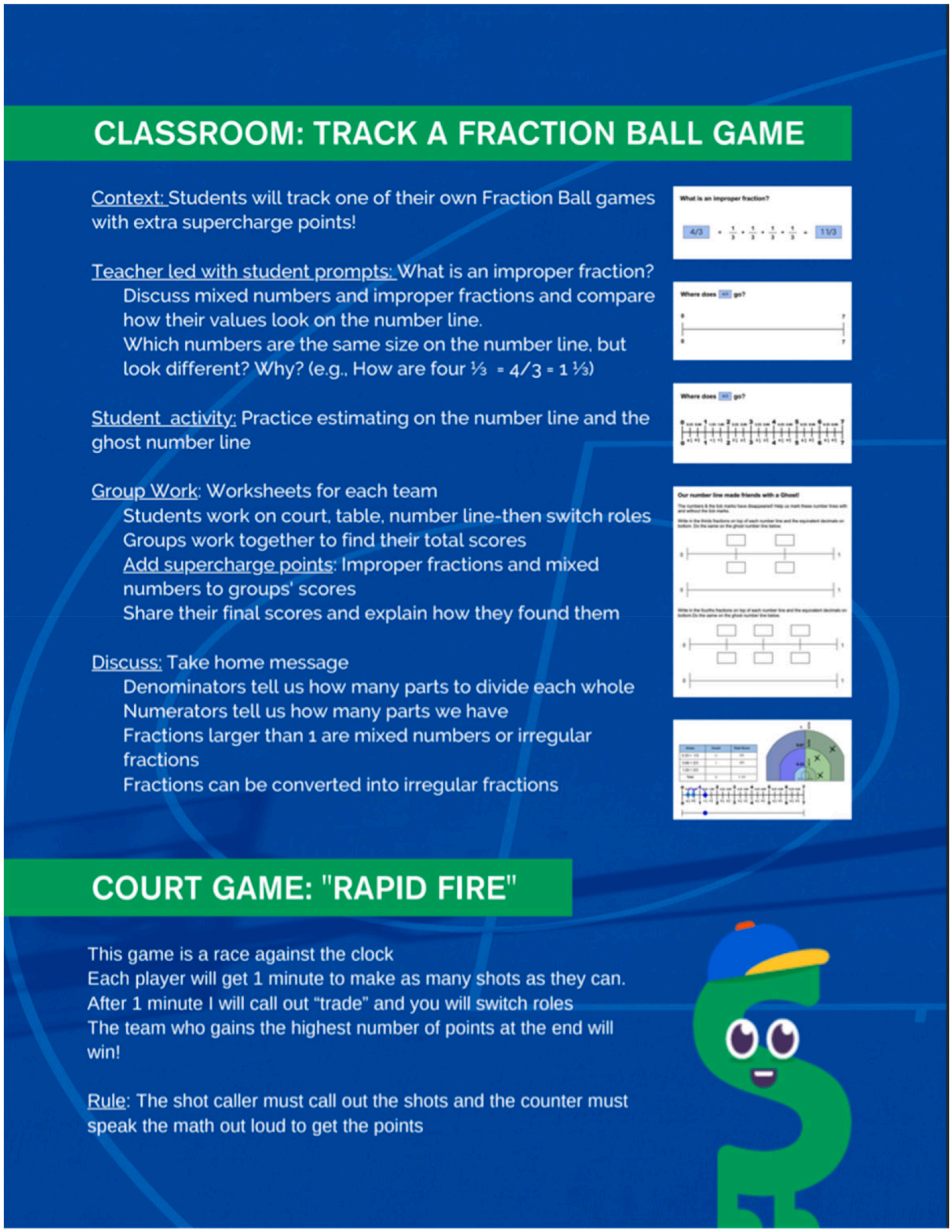
## MATERIALS

- Basketballs
- Pinnies
- Clipboard
- Tracking Worksheet

## STANDARD

**Math Standards: 4.NF.A.1, 4.NF.A.2, 4.NF.B.3, 4.NF.B.3, 4.NF.B.3.B, 4.NF.C.6, 4.NF.C.7, 5.NF.A.2**
**PE Standards: 4.NF.3, MP.1, MP.5**

.


CLASSROOM: TRACK A FRACTION BALL GAME
Context: Students will track one of their own Fraction Ball games with extra supercharge points!
Teacher led with student prompts: What is an improper fraction?
Discuss mixed numbers and improper fractions and compare how their values look on the number line.
Which numbers are the same size on the number line, but look different? Why? (e.g., How are four ⅓ = 4/3 = 1 ⅓)
Student activity: Practice estimating on the number line and the ghost number line
Group Work: Worksheets for each team
Students work on court, table, number line-then switch roles
Groups work together to find their total scores
Add supercharge points: Improper fractions and mixed numbers to groups' scores
Share their final scores and explain how they found them
Discuss: Take home message
Denominators tell us how many parts to divide each whole
Numerators tell us how many parts we have
Fractions larger than 1 are mixed numbers or irregular fractions
Fractions can be converted into irregular fractions
What is an improper fraction?
4/3
11/3
Where does go?
Our number line made friends with a Ghost!
COURT GAME: "RAPID FIRE"
This game is a race against the clock
Each player will get 1 minute to make as many shots as they can.
After 1 minute I will call out "trade" and you will switch roles
The team who gains the highest number of points at the end will win!
Rule: The shot caller must call out the shots and the counter must speak the math out loud to get the points

.

# Day 6

## BIG IDEA

Students will sum up their points made throughout the games (e.g. Make it Count, Rapid Fire) and analyze the strategies that they utilized during their game play. Students will express magnitude understanding of rational number vocabulary (e.g., In this game, students made one shot from the 3/3 line, and 2 from the $\frac{1}{3}$ line which equals $1\frac{2}{3}$) and magnitude comparison of their strategies used (e.g., In Make it Count, it was easier for our team to shoot from the 3/4ths line but in Rapid Fire, it was a better strategy to shoot closer to maximize points).

## FRACTION BALL CONNECTIONS

Students will use the Fraction Ball tracking worksheet and reflect on their game play and check their work.

## MATERIALS

Basketballs

Pinnies

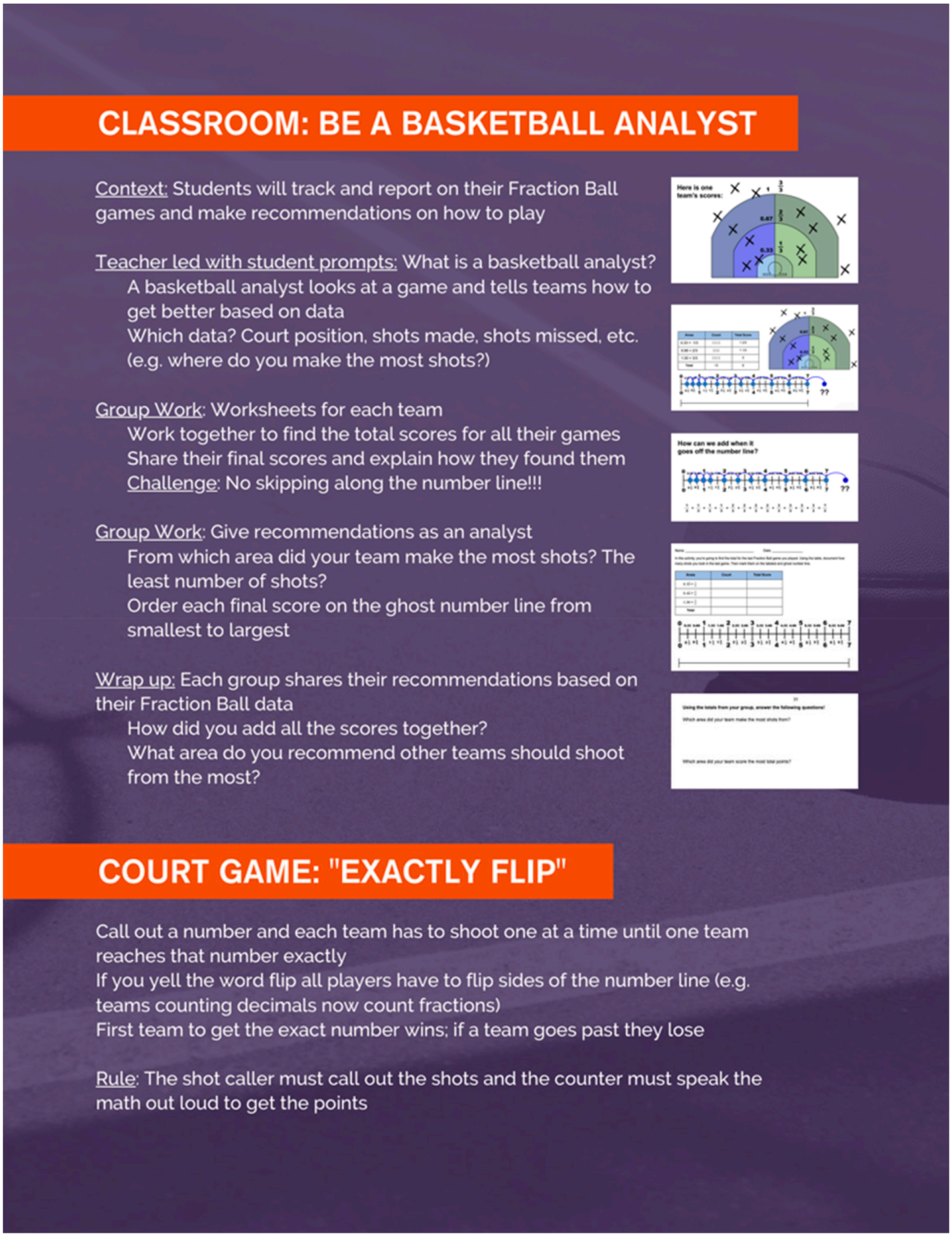
Clipboard

Tracking Worksheet

## STANDARDS

**Math Standards: 4.NF.A.1, 4.NF.A.2, 4.NF.B.3, 4.NF.B.3, 4.NF.B.3.B, 4.NF.C.6, 4.NF.C.7, 5.NF.A.2**
**PE Standards: 4.NF.3, 4.NF.4, MP.2, MP.4, MP.5, MP.6**

.


CLASSROOM: BE A BASKETBALL ANALYST
Context: Students will track and report on their Fraction Ball games and make recommendations on how to play
Teacher led with student prompts: What is a basketball analyst?
A basketball analyst looks at a game and tells teams how to get better based on data
Which data? Court position, shots made, shots missed, etc. (e.g. where do you make the most shots?)
Group Work: Worksheets for each team
Work together to find the total scores for all their games
Share their final scores and explain how they found them
Challenge: No skipping along the number line!!!
Group Work: Give recommendations as an analyst
From which area did your team make the most shots? The least number of shots?
Order each final score on the ghost number line from smallest to largest
Wrap up: Each group shares their recommendations based on their Fraction Ball data
How did you add all the scores together?
What area do you recommend other teams should shoot from the most?
Here is one team's scores:
How can we add when it goes off the number line?
COURT GAME: "EXACTLY FLIP"
Call out a number and each team has to shoot one at a time until one team reaches that number exactly
If you yell the word flip all players have to flip sides of the number line (e.g. teams counting decimals now count fractions)
First team to get the exact number wins; if a team goes past they lose
Rule: The shot caller must call out the shots and the counter must speak the math out loud to get the points

.

# Game Rules

## Rapid Fire

This game is a race against the clock. Each player will get 1 minute to make as many shots as they can. After 1 minute the coach will call out "trade" and you will switch jobs: the shooter becomes the rebounder, the rebounder becomes a counter, one counter stays, and the second counter becomes the shooter. After all 5 team members have a chance to shoot, the round will be over and the team with the highest amount of points wins!
4.NF.3, MP.1, MP.5

## Make It Count

In this game, each team is allowed to take exactly 15 shots from any distance they want. Bad shot, good shot, whatever happens, you only get 12 shots. Each player on each team gets 3 shots before you switch jobs. The goal is to get the highest amount of points possible. The team that has the highest score after 12 shots wins. In "Make it Count" you have to take turns shooting with the other team.
4.NF.3, 4.MD.2, 4.NF.5, 4.NF.7, & MP.2, MP.4, MP.5, MP.6,

## Exactly

In this game the coach will call out a number and each team has to alternate shooting baskets until one team reaches that number EXACTLY. This game is a little tricky–it's not about who makes the most shots, but which team gets to the EXACT number the coach called out the fastest. The first team to get to the exact number the teacher call out wins. But be very careful, if you go past the exact number you automatically lose that round!
4.NF.3, 4.NF.4, MP.2, MP.4, MP.5, MP.6

.

# Game Rules

## Rapid Fire "Target"

Each player will get 1 minute to make as many shots as they can but this time each team has a target number that they are trying to reach. The team that reaches the target number first wins! And just like the game exactly you have to get exactly to the number and not go past it. Remember, the team that gets to the target score faster wins! But if you overshoot the target score then you automatically lose.
4.NF.3, 4.NF.4, MP.2, MP.4, MP.5, MP.6

## Make It Count "Ghost":

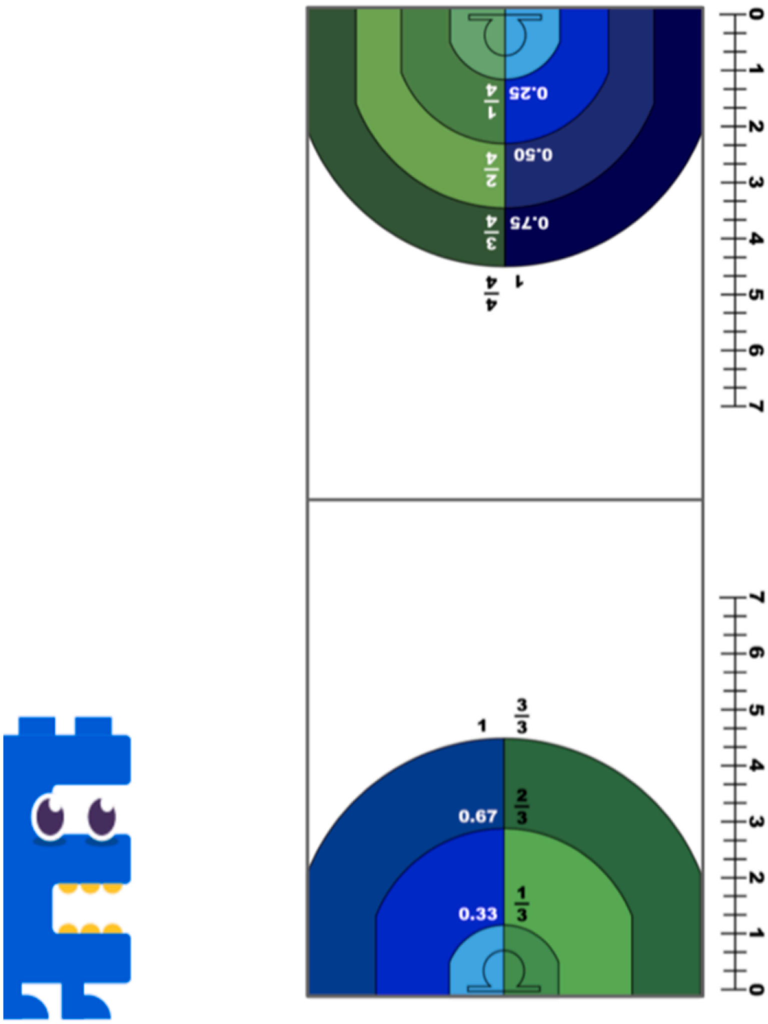
Here, the number line is drawn with chalk and has no numbers on it! Teams have to keep track of their score by estimating on the blank number line. As we all know, ghosts are very tricky, so the winner is not the team that has the most points at the end of 12 shots, but the winner is which team can most correctly estimate their score on the number line!
4.NF.3, 4.NF.4, MP.2, MP.4, MP.5, MP.6

## Exactly "Flip"

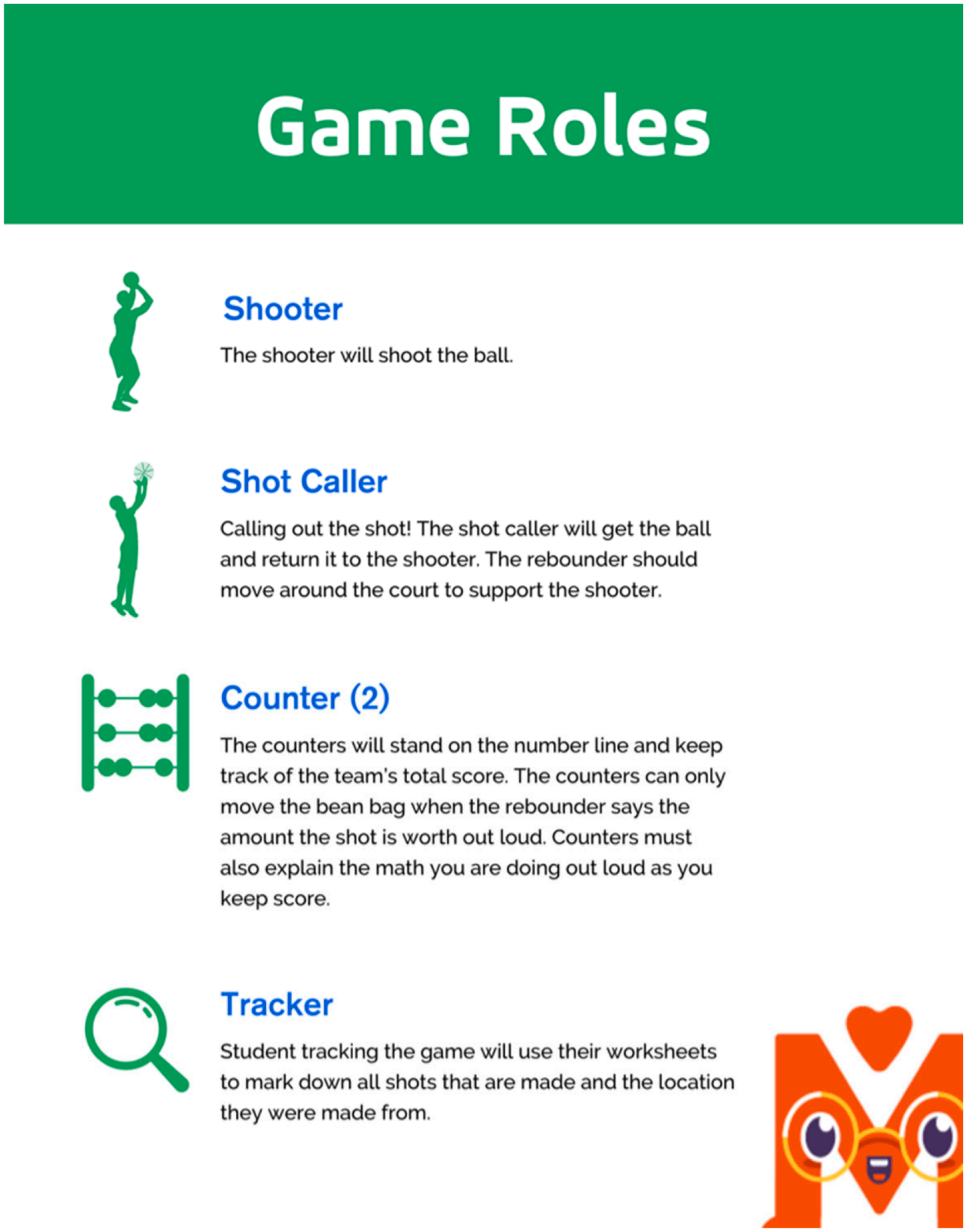
Like the game “Exactly”, but at any point during the game if I yell the word “flip” all players have to flip sides from the fraction side of the court to the decimal side. If you are shooting on the quarters side, you flip to the .25 side and the same on the thirds and .33 end of the court. If you are counting on the quarters number line you flip over to the .25 number line and the same on the thirds/ .33 end of the court. The rest of the rules for the game stay the same!
4.NF.5, 4.NF.7 4.NF.3, MP.2, MP.4, MP.5, MP.6

.

# Fraction Ball Court




.




.

## References

Alibali, M. W., & Knuth, E. J. (2018). Bridging psychology and mathematics education: Reflections on boundary crossing. *Journal of Numerical Cognition, 4*(1), 9–18.

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.

Begolli, K. N., & Richland, L. E. (2016). Teaching Mathematics by Comparison: Analog Visibility as a Double-Edged Sword. *Journal of Educational Psychology, 108*(2), 194–213. https://doi.org/10.1037/edu0000056

Begolli, K. N., & Richland, L. E. (2017). Bridging Cognitive Science and Real Classrooms: A Video Methodology for Experimental Research. *The Journal of Experimental Education*. https://doi.org/10.1080/00220973.2017.1347775

Begolli, K. N., Richland, L. E., Jaeggi, S. M., Lyons, E. M., Klostermann, E. C., & Matlen, B. J. (2018). Executive Function in Learning Mathematics by Comparing Representations: Incorporating Everyday Classrooms into the Science of Learning. *Thinking & Reasoning, 1–34*. https://doi.org/10.1080/13546783.2018.1429306

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*(1), 189.

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. (2021). Putting fractions together. *Journal of Educational Psychology, 113*(3), 556.

Brown, A. L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in-classroom Settings. *Journal of the Learning Sciences, 2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Bulus, M., Dong, N., Kelcey, B., & Spybrook, J. (2021). PowerUpR: Power Analysis Tools for Multilevel Randomized Experiments. *R package version, 1*(1). https://CRAN.R-project.org/package=PowerUpR.

Bustamante, A. S., Begolli, K., Alvarez-Vargas, D., Bailey, D., & Richland, L. (2022). *Fraction Ball: Playful and Physically Active Fraction and Decimal Learning.*. https://doi.org/10.1037/edu0000714

Carvalho, R., & da Ponte, J. P. (2017). Mental computation with rational numbers: Students' mental representations. *Journal of Mathematics Education, 10*(2), 17–29.

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

Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher, 32*(1), 5–8. https://doi.org/10.3102/0013189X032001005

Every student succeeds act, S. 1177—114th Congress (2015). https://www.gpo.gov/fdsys/pkg/BILLS-114s1177enr/pdf/BILLS-114s1177enr.pdf.

Fazio, L. K., Kennedy, C. A., & Siegler, R. S. (2016). Improving children's knowledge of fraction magnitudes. *PloS One, 11*(10).

Fuchs, L. S., Malone, A. S., Preacher, K. J., Fuchs, D., Wang, A. Y., & Pachmayr, R. (2019). Effects of fourth-and fifth-grade Super Solvers intervention on fraction magnitude understanding and calculation skill. Unpublished manuscript. http://frg.vkcsites.org/wpcontent/uploads/2019/07/Effects-of-Fourth-and-Fifth-Grade-Super-Solvers-Interventionon-Fraction-Magnitude-Understanding-and-Calculation-Skill.pdf.

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., Jordan, N. C., Siegler, R., Gersten, R., & Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology, 105*(3), 683–700. https://doi.org/10.1037/a0032446

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.

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The Effects of Feedback During Exploratory Mathematics Problem Solving: Prior Knowledge Matters. *Journal of Educational Psychology*. https://doi.org/10.1037/a0028389

Gentner, D. (1983). Structure-Mapping: A Theoretical Framework for Analogy*. *Cognitive Science, 7*(2), 155–170. https://doi.org/10.1016/S0364-0213(83)80009-3

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences, 14*(1), 69–110.

Hanington, B., & Martin, B. (2019). *Universal methods of design expanded and revised: 125 Ways to research complex problems, develop innovative ideas, and design effective solutions*. Rockport publishers.

Honig, M. I., & Coburn, C. (2008). Evidence-based decision making in school district central offices: Toward a policy and research agenda. *Educational policy, 22*(4), 578–608.

IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F.. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour, 4*(11), 1092–1094. https://doi.org/10.1038/s41562-020-00990-w

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional Complexity and the Science to Constrain It. *Science*, *342*(22 november 2013), 935–937. DOI: 10.1126/science.1238056.

Kokkonen, T., & Schalk, L. (2021). One instructional sequence fits all? A conceptual analysis of the applicability of concreteness fading in mathematics, physics, chemistry, and biology education. *Educational Psychology Review, 33*(3), 797–821.

Lawrence, L., Garcia, D. I., Ochoa, K., Valdez, S., Begolli, K., Bustamante, A. S., & Ahn, J. (2023). *Factors That Impact the Implementation of a Game Based Curriculum and Adaptations Teachers Design to Address Them.* Proceedings of the 17th International Conference of the Learning Sciences-ICLS 2023, pp. 918-921.

Lewis, N. A., Jr. (2021). What counts as good science? How the battle for methodological legitimacy affects public psychology. *American Psychologist, 76*, 1323–1333. https://doi.org/10.1037/amp0000870

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher, 48*(3), 158–166.

McCandliss, B. D., Kalchman, M., & Bryant, P. E. (2003). Design Experiments and Laboratory Approaches to Learning: Steps Toward Collaborative Exchange. *Educational Researcher, 32*(1), 14–16. https://doi.org/10.3102/0013189X032001014

Merriam, S. B. (1998). Qualitative Research and Case Study Applications in Education. Revised and Expanded from" Case Study Research in Education." Jossey-Bass Publishers, 350 Sansome St, San Francisco, CA 94104.

Penuel, W. R., Fishman, B. J., Haugan Cheng, B., & Sabelli, N. (2011). Organizing Research and Development at the Intersection of Learning, Implementation, and Design. *Educational Researcher, 40*(7), 331–337. https://doi.org/10.3102/0013189X11421826

Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and Practice in Technology Enhanced Learning, 2*(01), 51–74.

Polikoff, M., & Conaway, C. (2018). *Getting beyond "did it work?": Proposing a new approach to integrate research and policy*. Brown Center Chalkboard.

Pomerance, L., Greenberg, J., & Walsh, K. (2016). *Learning about Learning: What Every New Teacher Needs to Know*. National Council on Teacher Quality.

Reardon, S. F., Fahle, E. M., Kalogrides, D., Podolsky, A., & Zárate, R. C. (2019). Gender achievement gaps in US school districts. *American Educational Research Journal, 56*(6), 2474–2508.

Richland, L. E., Begolli, K. N., Simms, N., Frausel, R. R., & Lyons, E. M. (2017). Supporting Mathematical Discussions: The Roles of Comparison and Cognitive Load. *Educational Psychology Review, 29*(1). 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

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science, 316*(5828), 1128.

Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review, 27*(4), 587–597.

Santa Ana Unified School District. (2022). *July 13)*. Quick facts: District Overview. SAUSD.US. Retrieved from https://www.sausd.us/Page/5.

Schneider, M., & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology: Human Perception and Performance, 36*(5), 1227.

Schoen, R. C., Anderson, D., Riddell, C. M., & Bauduin, C. (2018). Elementary Mathematics Student Assessment: Measuring the Performance of Grade 3, 4, and 5 Students in Number (Whole Numbers and Fractions), Operations, and Algebraic Thinking in Fall 2015. Research Report No. 2018-24. Online Submission.

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 Fifth 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.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Mifflin and Company: Houghton.

Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives, 8*(3), 144–150. https://doi.org/10.1111/cdep.12077

Siegler, R. S., Thompson, C., & a., & Schneider, M.. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*, 273–296. https://doi.org/10.1016/j.cogpsych.2011.03.001

Simonsmeier, B. A., Flaig, M., Deiglmayr, A., Schalk, L., & Schneider, M. (2022). Domain-specific prior knowledge and learning: A meta-analysis. *Educational Psychologist, 57*(1), 31–54.

Slavin, R. E., & Fashola, O. S. (1998). *Show Me the Evidence! Proven and Promising Programs for America's Schools*. Corwin Press Inc, A Sage Publications Company, 2455 Teller Road, Thousand Oaks, CA 91320 (paperback: ISBN-0-8039-6711-X, $21.95; clothbound: ISBN-0-8039-6710-1, $49.95.

Star, J. R., Rittle-Johnson, B., & Durkin, K. (2016). Comparison and Explanation of Multiple Strategies: One Example of a Small Step Forward for Improving Mathematics Education. *Policy Insights from the Behavioral and Brain Sciences*. https://doi.org/10.1177/2372732216655543

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

Wang, Y., & Siegler, R. S. (2013). Representations of and translation between common fractions and decimal fractions. *Chinese Science Bulletin, 58*(36), 4630–4640.

Welsh, R. O. (2021). Assessing the quality of education research through its relevance to practice: An integrative review of research-practice partnerships. *Review of Research in Education, 45*(1), 170–194.

Wenger, E. (1999). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press.

Yeager, D. S., Romero, C., Paunesku, D., Hulleman, C. S., Schneider, B., Hinojosa, C., & Dweck, C. S. (2016). Using design thinking to improve psychological interventions: The case of the growth mindset during the transition to high school. *Journal of educational psychology, 108*(3), 374.

Yeon Lee, S., Christiaans, E., & Robinson, K. A. (2024). Development in context: Comparing short-term trajectories of expectancy, task values, and costs in four university STEM courses. *Contemporary Educational Psychology, 102282*. https://doi.org/10.1016/j.cedpsych.2024.102282