RG RESEARCH GROUP D/B/A

Instructional Research Group

Impact of a Tier 2 Fractions Intervention on Fifth-Grade Students' Fractions Achievement: A Technical Report

2018

Madhavi Jayanthi

Russell Gersten

Samantha Spallone

Joseph Dimino

Robin Schumacher

Keith Smolkowski

Instructional Research Group

Karen Karp

Johns Hopkins

Kelly Haymond

Instructional Research Group

4281 KATELLA AVE, SUITE 205 • LOS ALAMITOS, CALIFORNIA 90720
PHONE (714) 826-9600 • FAX (714) 826-9610

www.inresg.org

This project was funded by the National Science Foundation (Grant No. DRL-1535214).

Jayanthi, M., Gersten, R., Spallone, S., Dimino, J., Schumacher, R., Smolkowski, K., Karp, K., & Haymond, K. (2018). *Impact of a tier 2 fractions intervention on fifth-grade students' fractions achievement: A technical report.* Los Alamitos, CA: Instructional Research Group.

Author Note

The authors would like to thank Drew Bailey and Francis (Skip) Fennell for their fresh perspectives and counsel throughout the implementation of the study. In addition, the authors would like to thank Pamela Foremski and Christopher Tran for assisting with the research and preparation of the report.

Table of Contents

1. Executive Summary

The goal of this randomized controlled trial was to evaluate the effectiveness of a small-group intervention in fractions for fifth-graders who are performing below grade level in mathematics. The impact of the fractions intervention was assessed on fifth-grade at-risk students' understanding of foundational fractions concepts and procedural competence with fractions.

For the fractions intervention, lessons from the *TransMath*® curriculum (Level 2; Woodward & Stroh, 2015) were modified to create 52 thirty-five-minute lessons focused only on fourth- and fifth-grade level fractions content that could be used in a small-group setting. Lessons were structured so that each included review, demonstrations of concepts, student-led problem solving, and individual practice.

The *TransMath* curriculum was selected as a platform to examine fifth-grade at-risk students' understanding of foundational fractions concepts and procedural competence with fractions because it provides a balance among understanding of concepts, procedural competence, and problem solving. The curriculum emphasizes consistent use of number lines to build foundational fractions understanding and procedural competence, as well as discussions to enhance understanding and word problems to expand problem-solving abilities.

In this rigorous large-scale RCT, a sample of 1,123 students from three school districts across two U.S. states were screened at the beginning of the school year using a fractions measure developed by the research team. Two hundred and five students who scored between the 15th and the 37th percentile on the screening measure and received parental consent to participate were randomly assigned to the intervention (n = 102) and comparison (n = 103) conditions.

Students in the fractions intervention condition received 35-minute tutoring sessions 3-4 times a week. Fraction instruction using the 52 *TransMath* lessons was provided by trained tutors. The comparison condition (n = 103) was business as usual instruction (i.e., core classroom fractions instruction, including intervention or support traditionally provided by the school).

Results from the final analytic sample of 186 students (87 in intervention, 99 in comparison) showed that the intervention group significantly outperformed the comparison group on all outcome measures, which included an array of assessments used to measure both student understanding of foundational fractions concepts as well as procedural competence with grade-level fraction material (Hedges' g = .66 to 1.08; p < .0001).

2. Introduction

According to most current thinking on interventions for struggling learners in mathematics (e.g., Fennell, 2011; Fuchs et al., 2013; Gersten et al., 2009; Gersten, Taylor, Keys, Rolfhus, & Newman-Gonchar, 2014; Gersten et al., 2015), successful mathematics interventions are primarily *preventative*: that is, as much as possible, they *proactively* teach grade-level content or missing foundational concepts in a small-group setting that allows for much more support than would be feasible in a class of 30 students. Most argue that *preventative* (*Tier 2*) interventions should also build relevant foundational mathematical knowledge (both mathematics concepts and procedural proficiencies) essential for understanding grade-level content. In addition, intervention content should be based on best contemporary thinking about the mathematical content to teach. A source of confusion and occasional contention is the extent to which interventions for struggling learners should also include open-ended problem-solving activities and/or allow students to solve problems in more than one manner and discuss their reasons for choosing the strategy they used. Perhaps with the exception of one or two programs (e.g., Smith, Cobb, Farran, Cordray, & Munter, 2013), most studies of mathematics interventions do not include such procedures.

In the interest of exploring approaches that are more open-ended and reflect current thinking in mathematics education (Carpenter, Fennema, Franke, Levi, & Empson, 2014; Clarke, Roche, Sullivan, & Cheeseman, 2014) about the importance of not only building understanding of mathematical ideas and proficiencies, but also developing students' ability to solve problems and occasionally devise or invent their own strategies (as reflected for example in the Standards of Mathematical Practice MP1; National Governors Association Center for Best Practices [NGA Center], & Council of Chief State School Officers [CCSSO], 2010), the research team was interested in evaluating an intervention program that is consistent with these principles but still devotes a reasonable portion of each session to systematically building foundational mathematical skills and understandings. The research team selected the *TransMath*® fractions intervention (Woodward & Stroh, 2015) because it incorporates this balance among understanding of concepts, procedural competence, and problem solving. The program includes important fractions concepts and ideas articulated in the CCSS-M for Grades 4 and 5, as well as essential material in Grade 5 standards linking computational procedures to the underlying fractions concepts and ideas. Thus, the *TransMath* intervention fills the gaps in foundational fractions knowledge (e.g., understanding fractions as part of a set as well as part of a whole; locating fractions on a number line; equivalence) and proactively supports student learning of grade-level content in their regular classroom (i.e., fraction computation and linkage of computational methods to appropriate visual representations). *TransMath* also devotes significant time for students to solve problems and discuss their solutions.

The goal of this efficacy study was to conduct a rigorous randomized controlled trial to evaluate the effectiveness of *TransMath*, a small-group intervention in fractions for fifth-graders who are performing well below grade level in mathematics. The impact of the *TransMath* fractions intervention was assessed on fifth-grade at-risk students' (a) understanding of foundational fractions concepts, and (b) procedural competence with fractions. A wide range of measures was used to assess the impact, including two measures developed and tested by the IES Center for Improving Learning of Fractions: a fractions concepts measure aligned with Grade 4 CCSS-M and a test of fractions procedures aligned with Grades 4 and 5 CCSS-M. The impact was also assessed using a series of performance assessment tasks and number line estimation tasks. In addition, to better understand the learning environment that led to enhanced outcomes for this group of students, the *CLASS* observational system (Pianta, Hamre, & Mintz, 2012a) and the survey on instructional practices were used to measure the nature of the instruction in both intervention and comparison conditions.

3. Method

3.1 Setting and Participants

Student sample. One hundred and eighty-six fifth-grade students (89 boys, 97 girls) from three school districts (two urban districts on the West Coast and one from an urban-adjacent district in the Southeast) participated in this randomized controlled trial study. Student baseline characteristics are summarized in Table 1. Chi-square analysis and *t*-tests revealed no statistically significant differences between intervention and comparison students on any of the demographic variables or the pretest measures.

None of the 186 students had an IEP in mathematics. However, teachers confirmed that all students who met criteria had experienced persistent struggles learning mathematics, as one would expect given the screening and selection procedure (see the Screening criteria section, below). Thus, this would be considered a preventative intervention for students who did not master fourth-grade material in fractions, but whose problems were not so severe that they would be considered as students with mathematics learning disabilities or in need of one-on-one intensive intervention.

Table 1

Baseline Characteristics of the Student Analytic Sample

	Intervention (n = 87)	Comparison (n = 99)		
	Percent	Percent	χ^2 *(df)*	*p*
Gender			0.49 (1)	.485
Female	49.43	54.55		
Race/Ethnicity			3.41 (5)	.636
African American/Black	14.94	15.15		
Asian	5.75	6.06		
Hispanic/Latino	18.39	16.16		
White	42.53	34.34		
Multiracial	18.39	27.27		
Missing	0.00	1.01		
Free/Reduced Lunch			0.19 (2)	.909
Yes	57.47	54.55		
No	27.59	30.30		
Missing	14.94	15.15		
IEP/504			0.10 (2)	.950
Yes	6.90	8.08		
No	43.68	42.42		
Missing	49.43	49.49		
Pretest Measures	Mean Raw Score (SD)	Mean Raw Score (SD)	*t (df)*	*p*
TUF-4	11.49 (1.44)	11.45 (1.55)	0.18 (184)	.857
WRAT-4	97.91 (10.46)	96.18 (11.21)	1.08 (184)	.281
Procedures (Add/Sub)	5.69 (4.03)	5.13 (4.12)	0.93 (184)	.353
NLE 0–1	72.50 (10.54)	74.11 (12.14)	0.96 (184)	.339

Note. Total analytic sample N = 186 students (I = 87, C = 99). *WRAT-4* scores are standard scores. *WRAT-4* standard score of 97.91 = 44.45th percentile. *WRAT-4* standard score of 96.18 = 39.95th percentile.

***Screening criteria*.** A sample of 1,123 students from three school districts was screened at the beginning of the school year using a fractions measure developed by the research team, *Test for Understanding of Fractions, Fourth-Grade* (*TUF-4*; Instructional Research Group [IRG], 2014). The goal was to include students who score between the 15th and 35th percentile on a fractions measure. The lowest-35-percent cutoff has commonly been used in other studies examining the effectiveness of interventions for low-performing students (Fuchs et al., 2016; Powell & Fuchs, 2015).

The intent was to provide a preventative Tier 2 intervention, and therefore we excluded students below the 15th percentile as they might not benefit from a Tier 2 intervention alone and were likely to require a more intensive Tier 3 level of intervention.

Three hundred and twenty-six students (102 from District 1, 147 from District 2, and 77 from District 3) fell between the 15th and 37th percentile on a distribution for the *TUF-4* measure that was computed across the entire sample from the three districts.[1] Of the 326 students who qualified, 121 (31 from District 1, 50 from District 2, and 40 from District 3) were excluded prior to random assignment. Of these 121 students, 58 students were excluded by classroom teachers for a variety of reasons (e.g., conflicting schedules resulting from other services, not in need of intervention, no parental consent), and another 63 students were randomly excluded by the research team to make numbers more manageable at some school sites. This resulted in 205 students (71 from District 1, 97 from District 2, and 37 from District 3) who met criteria and had received parental consent to participate. These 205 students were randomly assigned to intervention (n = 102) and comparison (n = 103) conditions.

***Sample attrition*.** Students without posttests were counted toward attrition. A total of 17 students dropped out of the study for the following reasons: mobility (4 students: 3 intervention, 1 comparison), scheduling conflicts (6 students: 3 intervention, 3 comparison), parental consent rescinded (7 students from intervention). Of the remaining 188 students, two students were missing some pretest data. Thus, the final analytic sample (i.e., students with both pre- and posttest data) was 186 students (87 in intervention, 99 in comparison). Overall attrition was 9.27%, and differential attrition was

[1]Of the 186 students in the analytic sample, 30.65% scored above the 32nd percentile on the *TUF-4* screening measure.

10.82%. These are considered low by WWC Standards (U.S. Department of Education [U.S. ED], Institute of Education Sciences [IES], & What Works Clearinghouse [WWC], 2017).

Tutor sample. Ten tutors specifically hired by the research team participated in the study and provided intervention in mathematics. While none of the 10 tutors were employees of the school districts in which the study was conducted, they could be considered to be representative of those typically hired by districts to provide intervention within a response to intervention (RtI) framework. The tutor demographic data are summarized in Table 2. All were female; half possessed a master's degree, and half had prior experience tutoring in mathematics. Sixty percent were credentialed teachers. Average years of classroom teaching experience in elementary mathematics was 7.7 (*SD* = 8.92), and average experience teaching fifth-grade mathematics was 2.4 years (*SD* = 4.86).

Table 2

Baseline Characteristics of the Tutor Sample

Education Level	Percentage of Tutors
Bachelors	50
Masters	50
Experience Tutoring in Math	50
Years Teaching Elementary Math	
None	30
2–6	30
10–15	30
28	10
Years Teaching Fifth-Grade Math	
None	70
1	10
10+	20
Type of Teaching Credential[a]	
K–6	10
K–8	30
Multiple Subject K–12	30
None	40

Note. Total number of tutors in the analytic sample = 10.
[a]Percentages do not sum to 100 because one tutor has both K–8 and Multiple Subject K–12 credentials.

3.2 Study Design

In this multi-site randomized controlled trial, students were randomly assigned, by teacher, to intervention and comparison conditions. Randomization eliminates selection bias, within-school assignment leads to equivalence on district and school characteristics between the two conditions, and random assignment of students by teacher leads to equivalence in quality of core classroom mathematics instruction.

Core classroom mathematics instruction. Students in both intervention and comparison conditions received Tier 1 core classroom mathematics instruction from their mathematics teacher. *My Math* (McGraw-Hill Education, 2017b) was used as the core curriculum in District 1, *California Math* (McGraw-Hill Education, 2017a) was used in District 2, and *GO Math!* (Houghton Mifflin Company, 2018) was used in District 3.

Additional information on the nature of fractions instruction in Tier 1 core mathematics classrooms was obtained via classroom teacher surveys. Findings from these surveys are presented in Section 4.5.

Fractions intervention condition. Fraction instruction started in the last week of October 2016 and was completed by March 2017. Due to scheduling conflicts in four schools, the sessions did not end until April 2017. Based on local needs and schedules, the fractions intervention was implemented three times a week in nine tutoring groups and four times a week in 12 tutoring groups. At the start of the study, each tutoring group included 4–5 students (median group size = 5).

Fifty-two selected lessons from the *TransMath*® curriculum (Level 2; Woodward & Stroh, 2015) were used for the intervention. The *TransMath* curriculum was selected because it attempts to provide a balance between teaching mathematical ideas and teaching procedural proficiency, as well as attempting to explicitly link the two. Number lines were used consistently to build foundational understanding in both the part–whole, but especially the more difficult measurement interpretations of fractions, as well as understanding of the four operations when applied to fractions. Word problems were also emphasized to enhance understanding and build problem-solving abilities.

The *TransMath* curriculum was designed for use in large-group settings such as lower track or double-dose classes for students who have struggled with mathematics. It is geared toward Grades 4–8, and it includes the full range of the mathematics content covered in those grade levels in alignment with the Common Core State Standards. Each lesson is intended to be 55–60 minutes in duration.

For the fractions intervention, lessons from *TransMath* were reorganized by two members of the research team to create 52 thirty-five-minute lessons focused only on fractions that could be used in a small-group setting. The team identified the fractions material from *TransMath* (Level 2) that addressed both fourth- and fifth-grade-level content, so that lessons or lesson segments devoted to measurement, geometry, and whole number operations were eliminated. Lessons were sometimes restructured so that each lesson included review, demonstrations of concepts, student-led problem solving, and individual practice in a 35-minute lesson. Lesson structure consisted of the following segments:

1. Review: Practice previous day’s skills or prerequisite skills for the day (5 minutes),
2. Strategic Explicit Instruction: Instruction in key concepts along with checks for understanding (10 minutes),
3. Guided Practice: Solving problems as a group or with a partner (10 minutes), and
4. Student Explanations: Students solve problems independently and provide explanations for their strategies to the small group (10 minutes).

Through out the lesson, students were provided with specific corrective feedback to decrease the occurance of misconceptions and subsequent mathematical errors. In addition, the vocabulary students needed to understand the lesson content was listed at the beginning of each lesson and specifically taught during the lesson. Vocabulary word walls were maintained as new vocabulary was introduced.

Accurate student explanations was a goal of the intervention and, therefore, an integral part of the *TransMath* curriculum. To that end, students were provided with supports to help build their capacity in providing mathematically correct explanations of their solution methods. For example, a multi-step prompt card that listed four steps for writing explanations was developed. Steps were categorized as “thinking” steps (e.g.,“What’s the problem asking me?” and “What did I do to solve it?”) and “writing” steps (e.g., “Write all the steps using mathematical vocabulary.” and “Write why the answer makes sense.”) To further assist students in writing explanations that include mathematical language, each student was provided with a vocabulary card which listed relevant math terminology learned during the intervention (e.g.,denominator, equivalent, etc.). Thus, the multi-step prompt card and the vocabulary card were used during the fractions intervention to support students as they practiced explaining their thinking.

In addition, a prompt card with the acronmym LAPS was used to help students when they added mixed number fractions. The four steps prompted student to: Look (Are common denominators needed? Is grouping needed?), Alter (change denominators; regroup), Perform (add or subtract) and Simplify (reduce; regroup).

***Scope of the fractions content in the intervention*.** The intervention lessons covered the fractions content identified in contemporary mathematics standards for Grades 4 and 5 (e.g., Common Core State Standards in Mathematics [CCSS-M]; NGA Center & CCSSO, 2010; California Common Core State Standards for Mathematics [California Department of Education, 2013]). In general, fourth-grade standards focus on foundational fractions concepts such as equivalence and ordering (CCSS-M 4.NF.A.1) and understanding unit fractions (CCSS-M 4.NF.B.3). They also address addition and subtraction with like denominators (e.g., CCSS-M 4.NF.B.3.D) and multiplication of a fraction with a whole number (CCSS-M 4.NF.B.4). Fifth-grade standards extend fraction learning to addition and subtraction with unlike denominators (CCSS-M 5.NF.A.1), multiplication of a fraction by a fraction (CCSS-M 5.NF.B.4), and division of a whole number by a unit fraction (CCSS-M 5.NF.B.7).

Lessons 1–18 addressed foundational understanding. These lessons emphasized understanding what a fraction is, magnitude of fractions, equivalent fractions, developing understanding by comparing two fractions, ordering fractions from least to greatest, and estimating fraction placement on the number line. Lessons 19–28 focused on addition and subtraction of fractions, while Lessons 29–42 addressed multiplication and division of fractions. For instance, the lessons focused on the underlying concepts and procedures for (a) addition and subtraction of fractions with like denominators ($\frac{1}{2}+\frac{1}{2}$) and unlike denominators ($\frac{1}{2}-\frac{1}{3}$), (b) multiplication of a whole number times a fraction ($2 \times \frac{3}{4}$) and a fraction times a fraction ($\frac{1}{2} \times \frac{3}{4}$), and (c) division of a whole number divided by a unit fraction ($2 \div \frac{1}{4}$) and a unit fraction divided by whole number ($\frac{1}{4} \div 2$). The lessons also focused on critical fractions concepts related to computational procedures. Lessons explored, for instance, why fractions with unlike denominators cannot be added or subtracted before the problem is modified to include like denominators, or how the multiplication of two fractions involves finding a fraction of a fraction (e.g., $\frac{1}{2} \times \frac{4}{5}$ is the same as $\frac{1}{2}$ *of* $\frac{4}{5}$). The final set of lessons—Lessons 43–52—included material on adding and subtracting mixed numbers. Thus, lessons focused on fractions less than 1, fractions greater than or equal to 1, and mixed numbers. Throughout, other requisite skills with whole numbers were included (e.g., multiples and factors) to support solving fractions computation problems.

In most instances, the *TransMath* curriculum limited fractions that students encountered to those with familiar denominators (e.g., $\frac{1}{5}$, $\frac{1}{6}$, $\frac{2}{3}$ rather than $\frac{1}{11}$, $\frac{1}{13}$, $\frac{10}{15}$) that were easier to manipulate and understand. This helped students focus on the concepts being taught instead of getting distracted with intricate calculations with numbers that are rarely encountered.

Use of concrete and semi-abstract visual representations. While number lines were used as the central visual representation, Cuisenaire Rods® (a concrete manipulative) and area models were also used to visually represent fractions and serve as a means for demonstrating important concepts related to fractions. Cuisenaire Rods are linear, 3D, color-coded, and of various sizes, and can be used interchangeably to represent one whole or parts of a whole, thus allowing for hands-on exploration of fractions principles. The representations were included strategically, as the ultimate goal was for students to solve problems without needing concrete or other visual representations. Cuisenaire Rods were particularly useful in scaffolding student learning of fractions concepts and procedures. Area models were used occasionally to clarify fractions concepts (e.g., part–whole relationship, multiplication).

Refinements based on the pilot study. The fractions intervention was tested via a small-scale pilot study and revised based on the formative data. For additional information on the pilot study and revisions made to the fractions intervention, see Schumacher et al. (2018).

Comparison condition. The comparison condition was business as usual instruction. As not all the schools in the study were set up to provide formally structured Tier 2 mathematics intervention, we anticipated that only some of the students in the comparison condition would receive some form of intervention, with the rest receiving core classroom instruction, be it in mathematics, another academic subject (e.g., science), or a non-academic class like P.E. To determine how the class time was being spent by the comparison students during the 35-minute intervention block when intervention students were receiving the fractions intervention, surveys were administered to classroom teachers in fall and spring. Data from these surveys are discussed in Section 4.5.

3.3 Training of Tutors

Tutors attended a two-day training conducted by the research team. The first day of training began with a brief discussion of the purpose of the research project, followed by an overview of the topics covered in the 52 *TransMath* lessons. It concluded with demonstration and practice with concrete representations (Cuisenaire Rods, which were new to many of the tutors), as well as visual representations (number lines). The second day focused on techniques for facilitating verbal and written explanations (many developed during the pilot study; Schumacher et al., 2018), demonstrations of correct fraction procedures (e.g., how to regroup fractions, distinguishing between least common multiple and greatest common factor), and identifying and solving word problems that required the four operations. During the training, tutors were told that fidelity of implementation would be assessed. Tutors were issued a digital recording device and taught how to record their *TransMath* lessons and upload them onto a secure, password-protected website.

Ongoing support and feedback to tutors. All tutors received coaching from the research staff periodically throughout the course of the study to review lesson goals, address questions and concerns, and discuss their experiences implementing the curriculum (e.g., which aspects were difficult, which aspects were easy). Coaching phone calls were held every week for the first three weeks of the intervention with the entire group of tutors broken into two groups. A member of the research team listened to the audio recording of each tutor’s third or fourth lesson and provided them with immediate feedback. If the tutor was deemed to be “weak,” additional checks of the audio recordings were conducted, and feedback was provided. As the intervention progressed, coaching calls were held every three weeks with the whole group and on an as-needed basis with individual tutors.

3.4 Measures

Student measures. The following student measures were used in the study.

1. *Test for Understanding of Fractions, Fourth-Grade (TUF-4)*. The *TUF-4* (IRG, 2014) was administered at pretest. The measure was used (a) to screen students for eligibility, (b) as a potential covariate, (c) as a means to help describe pretest-posttest growth in fractions knowledge for the two samples, and (d) as one of the major posttest measures. The measure consists of 26 multiple-choice fraction items based on third- and primarily fourth-grade CCSS in mathematics. The items were selected from publicly available NAEP assessments for Grades 4 and 8 (National Center for Education Statistics, 2009), *Illustrative Mathematics* (Illustrative Mathematics, 2013), and the various measures used by researchers from the IES fraction item bank at the Center for Improving Learning of Fractions (e.g., Fuchs, et al., 2013; Jordan et al., 2013). This item pool was then subjected to review by two research mathematicians involved in mathematics education, Kristin Umland and Jim Lewis, who scrutinized the accuracy of the mathematical language used and the extent to which understanding of key mathematical ideas and concepts was assessed. They also assisted in ensuring that all relevant Grade 4 (and to some extent Grade 3) CCSS-M addressing fractions were addressed. The measure demonstrated a coefficient alpha reliability of .80 in the current study and .86 in a previous, larger-scale study.[2] Item difficulty estimates range from

[2]The Cronbach's alpha based on sample of 5,005 Grade 4 students is 0.86 (Jayanthi, Gersten, Taylor, Smolkowski, & Dimino, 2017).

−14.26 to 3.82. Overall, 11 of the 26 items have difficulty estimates below zero. Discrimination estimates for the 26 items range from −2.23 to 0.76.

2. *Wide Range Achievement Test 4 (WRAT-4): Math Computation subtest*. The 40-item Math Computation subtest is a measure of general mathematics achievement. It was administered at pretest only as a potential covariate, because the *TUF-4*, by design, would have a restricted range, since it served as a screener. The *WRAT-4* is a brief overview of general mathematics proficiency. Median reliabilities range from 0.83 to 0.87 (Wilkinson & Robertson, 2006).

3. *Test for Understanding of Fractions, Fifth-Grade (TUF-5)*. Since the measure is aligned with Grade 5 CCSS-M related to fractions, *TUF-5* (IRG, 2015) was used at posttest only. The items were also reviewed by mathematics educators (Francis Fennell and Karen Karp) to ensure comprehensiveness and alignment with CCSS-M standards and precision of mathematical language. It includes 18 items derived from NAEP and PARCC fourth- and fifth-grade assessments (Pearson Education, 2015). The internal consistency for the measure is .76. Item difficulty estimates range from −9.87 to 3.39. Overall, 5 of the 18 items have difficulty estimates below zero. Discrimination estimates for the 18 items range from −0.10 to 3.52.

4. *Test of Fraction Procedures*. This 24-item measure, used as a posttest, was adapted from the measure developed by Jordan et al. (2013). It assesses students' skill with the four arithmetic operations involving fractions, as articulated in the Grade 5 CCSS-M standards. It is an adapted version of the measure. The measure has an internal consistency of .89. Item difficulty estimates range from −3.40 to 1.86. Overall, 11 of the 24 items have difficulty estimates below zero. Discrimination estimates for the 18 items range from 0.35 to 4.29. A shortened 12-item version of this test, containing

only the addition and subtraction items, was used as a pretest. The internal consistency for the pretest measure is .86.

5. *Number Line Estimation (NLE)*. Two measures, *NLE 0–1* (nine items; Fuchs et al., 2016) and *NLE 0–2* (19 items; Fuchs et al., 2016) were used to assess students' ability to place fractions on number lines with endpoints from 0 to 1 and 0 to 2, respectively. *NLE 0–1* was used as both a pretest and a posttest, whereas the more difficult *NLE 0–2* was used only as a posttest. Test-retest reliability for these measures is .80 (Fuchs et al., 2016).

6. *Curriculum-aligned assessment*. This 25-item measure is aligned with the intervention and includes multiple choice and constructed response items that assess student proficiency in the concepts and procedures covered by the intervention. Items on this test were selected from the end-of-unit tests and performance assessment tasks present in the *TransMath* curriculum. Cronbach's alpha for the 25-item scale is .78. Item difficulty estimates range from −2.55 to 3.65. Overall, 16 of the 25 items have difficulty estimates below zero. Discrimination estimates for the 25 items range from 0.18 to 5.26.

7. *Performance-based assessments*. The assessment battery also included five performance-based assessments that were administered throughout the course of the intervention to all participating intervention students and to a randomly selected sub-sample of 34 comparison students. Each assessment included one constructed response type item (Chval, Lannin, Jones, & Dougherty, 2013) and was a topic covered in the intervention. See Table 3 for a description of the items from each performance assessment. Students were asked not only to solve the problem, but also to provide a written explanation for their response (i.e., their rationale).

The research team developed a unique rubric to score each problem. The rubrics were developed in an iterative fashion by initially scoring a small set of student responses. Two members of the research team then used the final rubrics to score all the student responses. Each rubric assessed the accuracy of the answer and the quality of the written explanation. Full points were awarded for a correct and simplified answer, and partial points were given for a correct but unsimplified answer. The students' written explanations were graded according to a set of elements that were unique to each problem. The research team examined each problem and considered what critical understandings the students had to have for solving the problem (e.g., rationale for solving the problem, which operation to use and why, and how they would execute the operation). For example, the scoring rubric for the written explanations for the fifth performance assessment (i.e., Manny is training for a marathon. He ran $10\frac{2}{5}$ miles on Saturday. On Sunday, he ran $7\frac{9}{10}$ miles. How many more miles did Manny run on Saturday than Sunday?) allowed for a total of five points, one point for each of the following elements: (a) student writes about changing $\frac{2}{5}$ to $\frac{4}{10}$ or about common denominators, (b) student presents a rationale for subtracting, (c) student mentions that you cannot subtract without regrouping, (d) student explains the regrouping process, and (e) student mentions simplifying the answer. In addition, the scoring rubrics for performance assessments 3, 4, and 5 also awarded an additional point to students who identified and applied the correct operation to solve the problem.

Given the moderate level of inference necessary for scoring the assessments, the inter-rater agreement between the two scorers was defined as any scores that fell within one point of each other (i.e., plus or minus one). The mean inter-rater reliability was 99.42% for correct answer, 96.54% for appropriate written explanation, and 100%

for correct operation.

Table 3

Performance-based Assessments

Assessment	Problem
1	Mark where $\frac{2}{5}$ is located on the number line below. How did you know where to place $\frac{2}{5}$ on the number line? Explain your thinking.
2	Order the fractions from least to greatest on the blank spaces below. Then, mark and label these fractions on the number line: $\frac{1}{2}$ $\frac{3}{4}$ $\frac{1}{8}$ $\frac{7}{12}$. How did you know where to place each fraction? Explain your thinking.
3	Use a number line, draw pictures, or use numbers to solve the word problem. On Saturday, Jessie walked $\frac{1}{2}$ of a mile to the park. Then, from the park she walked $\frac{4}{10}$ of a mile to her friend Emily's house. Next, Jessie and Emily walked $\frac{7}{10}$ of a mile together. How far did Jessie walk on Saturday? Explain your thinking.
4	Solve the word problem. Use pictures, number lines, or numbers to show your problem solving. Bella likes to build with Legos. In her set of Legos, $\frac{1}{5}$ are red. Bella used $\frac{2}{3}$ of her red Legos to build a fire truck. What fraction of her total set of Legos did she use to build the fire truck? Explain your thinking.
5	Solve the word problem. Use pictures, number lines, or numbers to show your work. Simplify your answer by putting into lowest terms. Manny is training for a marathon. He ran $10\frac{2}{5}$ miles on Saturday. On Sunday, he ran $7\frac{9}{10}$ miles. How many more miles did Manny run on Saturday than Sunday? Explain your thinking.

Observations of tutoring groups. The *Classroom Assessment Scoring System (CLASS): Upper Elementary* (Pianta et al., 2012a) was used to assess the nature of intervention instruction in each tutoring group. The *CLASS* examines the quality of instruction by capturing the nature of the interactions between students and tutors. The *CLASS* consists of three domains—Emotional Support, Classroom Organization, and Instructional Support—and 12 dimensions that assess the quality of tutors' instructional and social interactions. These instructional dimensions are rated on a 7-point Likert Scale. See Table 4 for descriptions of each *CLASS* dimension.

Each tutoring group was observed during one intervention session by a trained and certified *CLASS* observer. Detailed information on the *CLASS* ratings across the 21 tutoring groups is presented in Section 4.2.

Four tutoring groups were observed by two observers to establish inter-rater reliability, which was calculated using the percent agreement formula. The raters were considered to be in agreement if their ratings were within one point of each other on the 7-point Likert Scale. Inter-rater reliability within one point difference was 100%.

Table 4

Upper Elementary CLASS Domains and Dimensions

Domain	Dimension	Indicators
Emotional Support	Positive Climate	Relationships; positive affect; positive communications; respect
	Teacher Sensitivity	Awareness; responsiveness to academic and social/emotional needs and cues; effectiveness in addressing problems; student comfort
	Regard for Student Perspectives	Flexibility and student focus; connections to current life; support for autonomy and leadership; meaningful peer interactions
Classroom Organization	Behavior Management	Clear expectations; proactive; effective redirection of misbehavior; student behavior
	Productivity	Maximizing learning time; routines; transitions; preparation
	Negative Climate	Negative affect; punitive control; disrespect
Instructional Support	Instructional Learning Formats	Learning targets/organization; variety of modalities, strategies, and materials; active facilitation; effective engagement
	Content Understanding	Depth of understanding; communication of concepts and procedures; background knowledge and misconceptions; transmission of content knowledge and procedures; opportunity for practice of procedures and skills
	Analysis and Inquiry	Facilitation of higher-order thinking; opportunities for novel application; metacognition
	Quality of Feedback	Feedback loops; scaffolding; building on student responses; encouragement and affirmation
	Instructional Dialogue	Cumulative content-driven exchanges; distributed talk; facilitation strategies
	Student Engagement	Active engagement

Surveys. The following surveys were administered.

***Classroom teacher survey of instructional practices*.** Classroom teachers were asked to complete a survey in fall and spring to determine when instruction in fractions was provided to intervention and comparison students in their core mathematics classrooms. Teachers were also asked several questions about the nature of their fractions instruction (e.g., fraction content that was covered, type of representations used, opportunities for students to explain their understanding). In addition, teachers were probed about the activities of the comparison group students during the 35 minutes when the intervention students were receiving the *TransMath* fractions intervention. Responses from the classroom teacher survey are summarized in Section 4.5.

Student, tutor, and classroom teacher appraisal surveys. Students, tutors, and classroom teachers were given appraisal surveys at the end of the intervention in spring to solicit their feedback on the perceived benefits of the fractions intervention. Findings from these appraisal surveys are presented in Section 4.5.

3.5 Fidelity of Implementation

The research team selected a purposeful sample of eight *TransMath* lessons to assess fidelity (Lessons 4, 11, 20, 27, 31, 35, 43, and 47). These lessons were selected as they cover all phases of the instructional period and address critical intervention topics and instructional approaches. Fidelity was assessed on lessons that included instruction in foundational fractions concepts (e.g., equivalence and magnitude), the four operations, addition and subtraction with regrouping, written explanations, and word problems. Fidelity lessons also sampled instruction that relied on manipulatives and the number line representation. All eight sessions were scored for both fidelity to the procedures and activities required by the intervention curriculum and perceived implementation quality.

All 52 *TransMath* lessons were audio recorded, and the tutors were not told which audio recordings would be checked for fidelity. For one tutoring group, fidelity was assessed through classroom observations (rather than by listening to audio recordings) as the parent/guardian of one participating student in the group declined permission to audio record.

Procedural fidelity was assessed using checklists developed by the research team based on the lessons' curricular content. The checklists included 55 procedures and activities, on average (range = 34 to 82 items). Each procedure/activity on the checklist was marked as observed or not observed. Fidelity was calculated as percentage of procedures implemented (number of procedures observed ÷ total number of procedures [observed and not observed] × 100).

Quality of implementation was assessed by rating the tutors on "qualities" generally associated with mathematics instruction: tutor's pacing of the lesson, clarity and mathematical correctness of language, providing specific math-oriented praise, ability to enhance student explanations, and ability to maintain a positive rapport with students. The fidelity raters (members of the research team) also rated the implementation quality based on their perception of the students' grasp of the content during the intervention session. They also provided an overall rating. All seven items were rated on a 5-point Likert scale, with 1 = low and 5 = high. See Section 4.4 for fidelity of implementation data.

To determine inter-rater reliability of the fidelity data, 22 randomly selected audio recordings (13% of the total recordings) were coded by two raters. Given the moderate level of inference necessary for coding the lessons, the agreement was defined as any ratings that fell within one point of each other. The mean inter-rater reliability was 81.57% for the procedural items and 75.97% for the quality of instruction items.

3.6 Data Analysis

Main impact analysis. For all outcomes, we conducted an analysis of covariance (ANCOVA) for partially nested data described by Bauer, Sterba, and Hallfors (2008) and Baldwin, Bauer, Stice, and Rohde (2011). The study design called for the randomization of individual students in each class to either the *TransMath* intervention condition, with students nested within small groups, or a non-nested comparison condition. Thus, *TransMath* groups, but not the unclustered comparison group, required consideration of a group-level variance at the tutoring group level. This required an analytic model that accounted for the potential heterogeneity of variance across conditions (Roberts & Roberts, 2005). We used Satterthwaite approximation to determine degrees of freedom.

Because the residual variances may have differed between conditions, we tested the assumption of homoscedasticity of residuals and reported results of the most appropriate model for each outcome measure. We tested whether the homoscedastic and heteroscedastic models could be assumed equivalent with a likelihood ratio test and reported the simpler model if we were able to accept the equivalence of the two models. Because this tests the simpler model's noninferiority when compared to the more complex model, we reversed the null and alternative hypotheses and, hence, the Type I and Type II error rates, α and β, common among equivalence or noninferiority trials (Dasgupta, Lawson, & Wilson, 2010). For this reason, and the poor statistical power associated with tests of variance structures (Kromrey & Dickinson, 1996), we set α = .20 as our Type I error rate and reported the more complex model unless we were relatively certain that the two were equivalent.

Sensitivity analyses. We conducted several variations on the analysis to test the sensitivity of the main-effect results to the analytic method. Due to the conceptual and technical flaws noted by some with covariate adjustment (e.g., Spector & Brannick, 2011; Willett, 1988), we examined condition differences on net gains with a mixed model Time × Condition analysis (Murray, 1998), modified to account for students partially nested within small groups (see Clarke et al., 2016, for details). This analysis included only a subset of the three measures with both pretest and posttest data. We also tested models without covariates, recommended by Simmons, Nelson, and Simonsohn (2011). We next tested a set of mixed-models that nested students and their small groups within teachers or schools. Finally, Baldwin, Murray, and Shadish (2005) and Roberts and Roberts (2005) have argued that for individuals nested within an interventionist, the analysis should treat the interventionist as the unit of analysis. Hence, we also examined the effects for students nested within tutor rather than small group, as some tutors taught multiple small groups.

Model estimation. We fit the aforementioned statistical models to our data using SAS PROC MIXED version 14.2 (SAS Institute, 2016) and restricted maximum likelihood estimation. The models assume independent and normally distributed residuals. We addressed the first assumption (van Belle, 2008) by explicitly modeling the multilevel nature of the data. Multilevel regression methods have also been shown to be quite robust to violations of normality (e.g., Hannan & Murray, 1996).

Effect sizes and multiple tests. To interpret results, we computed the Hedges' *g* effect size for each fixed effect of condition as recommended by the What Works

Clearinghouse (U.S. Department of Education et al., 2017). We also corrected for multiple tests with the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) and reported the original *p*-values as well as the Benjamini-Hochberg corrected *p*-values (p_{BHc}) for each outcome. We adjusted *p*-values separately for the seven tests of main effects.

Moderation analysis. To determine whether students responded differentially to the *TransMath* intervention based on initial math achievement on *WRAT-4*, *TUF-4*, *Test of Fraction Procedures* (addition and subtraction items only), or *NLE 0–1*, we expanded the ANCOVA to include each moderator and its interaction with condition. We also examined whether the impacts varied by district, gender, and free and reduced lunch status. The tests for equivalence between homoscedastic and heteroscedastic variances were conducted in the same manner as for main effects, described above.

Mediation analysis. For this analysis we examined the mediating effect of improved number line estimating skills on fraction understanding. We focused on the *Number Line Estimation* measure as a potential mediator as it is a widely used and accepted measure of students' grasp of the measurement interpretation of fractions (e.g., Siegler & Pyke, 2013; Siegler, Thompson, & Schneider, 2011) and was shown to mediate fractions achievement (Fuchs et al., 2016). We examined the mediating effect on *TUF-4*, since it was the only measure that was given as a pre- and post-test, thus allowing the calculation of a gain score. The use of gain scores decreases the biases associated with mediation tests discussed by Maxwell and colleagues (Mitchell & Maxwell, 2013). The primary concern is that mediation models without repeated assessments of the mediator and outcome measures could produce biased results (Judd & Kenny, 1981). Our approach using gains has fewer problems with biases than the standard model (von Soest & Hagtvet, 2011), but it can still only document correlations. Thus, these analyses are correlational, so neither temporal nor causal inferences are warranted.

To offer evidence in support of mediation, we tested whether the effect of (X) condition on (Y) gains on the *TUF-4* may have been explained by the indirect effect through (M) gains on the *NLE 0–1*. The indirect-effects model roughly follows Baron and Kenny's (1986) causal steps approach, which has considerable intuitive appeal but will, with cross-sectional and even sequential data, produce biased estimates (e.g., Judd & Kenny, 1981; Mitchell & Maxwell, 2013). As a potential remedy, von Soest and Hagtvet (2011) demonstrated an approach to mediation with growth curve models. With only two assessments, however, we estimated pre-post gains over time. The model does not use latent variables, so it does not benefit from increased power. Individual growth over time has been shown to be a "natural extension of the observed difference score" (Willett, 1988, p. 414), so the methods of von Soest and Hagtvet (2011), when applied to gains, should test whether the data are consistent with the hypothesis of mediation.

The indirect-effects model was estimated in Mplus (Muthén & Muthén, 1998-2017), and to address the nonnormality of the sampling distribution for the test of the indirect path from condition through *NLE 0–1* to *TUF-4*, we used bias-corrected bootstrapped confidence intervals (Preacher & Hayes, 2008) based on 5,000 samples. This analysis did not nest intervention students within small groups or include covariates. Our sensitivity analyses of impact estimates produced similar intervention effects and standard errors under these conditions.

4. Results

4.1 Impact of the Fractions Intervention on Student Fractions Achievement

Descriptive data. Unadjusted pretest and posttest means and standard deviations are listed in Table 5. Note that the sample of students selected to participate in the study was between the 15^{th} and the 37^{th} percentile on the *TUF-4* measure based on a distribution that was computed across the entire tested sample (1,123 students) from the three districts. Thus, while the selected students were performing well below average in the area of fractions, they were performing at a higher level (between the 40^{th} and the 44^{th} percentile) but still below average on *WRAT-4*, a nationally normed test of general mathematics achievement.

The mean percent correct on *TUF-5*, an 18-item contemporary grade-level measure of fractions achievement aligned with Grade 5 CCSS standards, was 46.45 for students who participated in the intervention and 32.55 for students who received business as usual. The mean percent correct on the 24-item *Test of Fraction Procedures* assessing fifth-grade level fraction computation in addition, subtraction, multiplication, and division was 55.04 for intervention students and 32.09 for comparison students.

Table 5

Descriptive Data on Student Pre- and Posttests

	Intervention (n = 87)		Comparison (n = 99)	
	Unadjusted		Unadjusted	
Pretest Measure	Mean Test Score[a] (SD)	Mean Percent (SD)	Mean Test Score[a] (SD)	Mean Percent (SD)
TUF-4 (Screener)	11.49 (1.44)	44.21 (5.53)	11.45 (1.55)	44.06 (5.95)
WRAT-4	97.9[b] (10.46)	n/a	96.18[c] (11.21)	n/a
Procedures (Add/Sub)	5.69 (4.03)	23.71 (16.80)	5.13 (4.12)	21.38 (17.18)
NLE 0–1	n/a	72.50 (10.54)	n/a	74.11 (12.14)
Posttest Measures				
TUF-4	16.84 (4.16)	64.77 (16.00)	13.46 (4.13)	51.79 (15.87)
TUF-5	8.36 (3.92)	46.45 (21.78)	5.86 (3.25)	32.55 (18.05)
Test of Fraction Procedures	26.42 (10.79)	55.04 (22.48)	15.40 (9.46)	32.09 (19.72)
Curriculum-aligned Measure	16.16 (4.66)	64.64 (18.64)	11.77 (3.55)	47.07 (14.19)
NLE 0–1	n/a	90.65 (7.53)	n/a	80.03 (11.67)
NLE 0–2	n/a	87.24 (8.33)	n/a	80.82 (8.15)

Note. Total sample size N = 186 students (I = 87, C = 99). Sample size for *TUF-5*, *Test of Fraction Procedures (Full)*, *NLE 0–1*, and *NLE 0–2* posttests is 185 students (I = 86, C = 99). [a]*WRAT-4* scores are standard scores; the rest are raw test scores. [b]*WRAT-4* standard score of 97.91 = 44.45th percentile. [c]*WRAT-4* standard score of 96.18 = 39.95th percentile. n/a = not available.

Impact on student posttests. The results of the partially nested analyses that compared intervention students in small groups to unclustered comparison students at posttest are presented in Table 6. The Hedges' g values range from .66 to 1.08, and p-values were all less than .0001, even after the Benjamini-Hochberg correction.

The findings provide an estimate of the fractions intervention's impact on students' understanding of foundational fractions concepts and procedural competence with fractions. Two fraction measures—*TUF-4* and *TUF-5*—were used to assess impacts on students' proficiency with foundational and grade-level fractions content. The effect size (Hedges' *g*) on the *TUF-4* measure was .78 and was statistically significant ($p < .0001$). The impact was also statistically significant for *TUF-5* ($g = .66$, $p < .0001$).

The *Test of Fraction Procedures* was used to measure student proficiency with grade-level computation in the areas of addition, subtraction, multiplication, and division of fractions. Analyses revealed that students who received the *TransMath* fractions intervention significantly outperformed students from the comparison group ($g = 1.07$; $p < .0001$). The *Curriculum-aligned Measure* was used to assess the intervention's impact on students' understanding of the fraction content covered in the *TransMath* intervention. The impact was statistically significant ($p < .0001$), with an effect size (*g*) of 1.06.

The *NLE 0–1* and *NLE 0–2* measures were used to determine if there were additional impacts on students' estimation of relevant fraction magnitude. Students in the intervention condition scored significantly better than those in the comparison group on the *NLE 0–1* ($g = 1.08$; $p < .0001$) and *NLE 0–2* ($g = .80$; $p < .0001$).

In Table 6, the top set of rows presents the fixed effects, followed by the variances in the next set of rows, with details about the tests of condition in the third set of rows. The bottom two rows of the table show the likelihood ratio test results that compared homoscedastic residuals with heteroscedastic residuals, and the tables report a different number of variances depending on the results. The data fit the homoscedastic model that assumed an equivalent residual variance across conditions for the *Curriculum-aligned Measure*, *TUF-4*, *TUF-5*, and *Procedures* measures. The data fit the heteroscedastic model ($p < .20$) for the *Procedures (Add/Sub)*, *NLE 0–1*, and

NLE 0–2 measures. Although the variance structures differed between these models, the condition effect estimates and their statistical significance values were very similar for the heteroscedastic and homoscedastic models. All analytic models included the *WRAT-4* and *NLE 0–1* pretest measures as covariates, which were statistically significant in each model.

See Table 7 for adjusted posttest means and unadjusted standard deviations.

Table 6

Results of a Partially Nested Mixed-Model Analysis of Covariance on Students' Posttests

		TUF-4	*TUF-5*	*Test of Fraction Procedures*	*Curriculum -aligned Measure*	*NLE 0–1*	*NLE 0–2*
Fixed Effects	Intercept	1.46 (2.48)	−5.50* (2.34)	−14.92* (6.15)	.94 (2.74)	63.77**** (6.22)	67.48**** (4.98)
	Condition (Intervention)	3.25**** (.56)	2.36**** (.46)	10.79**** (1.37)	4.33**** (.51)	10.79**** (1.20)	6.56**** (.95)
	WRAT-4 Pretest	.16**** (.02)	.14**** (.02)	.40**** (.06)	.14**** (.02)	.27**** (.06)	.23**** (.05)
	NLE 0–1 Pretest	.14**** (.02)	.09**** (.02)	.31**** (.05)	.11**** (.02)	−.37**** (.05)	−.35**** (.04)
Variances	Intervention Group Intercept	2.87* (1.31)	1.00 (.77)	16.95* (7.13)	.49 (.95)	4.88 (4.91)	.75 (3.77)
	Residual	6.44**** (1.11)	6.90**** (.96)	39.33**** (6.33)	10.50**** (1.41)		
	Intervention Residual					38.21**** (6.68)	44.57**** (7.69)
	Comparison Residual					68.03**** (12.03)	31.24**** (5.94)
ICC	Intervention Groups	.31	.13	.30	.04	.11	.02
Hedges' *g*	Condition	0.784	0.660	1.068	1.055	1.083	0.796
p-values	Condition	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
BH *p*-values	Condition	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
df	Condition	47	61	55	62	66	49
Likelihood ratio χ^2		0.34	0.30	0.01	0.90	5.08	2.74
p-values		.558	.585	.919	.342	.024	.098

Note. Total sample size *N* = 186 students (I = 87, C = 99). Sample size for *TUF-5*, *Test of Fraction Procedures*, *NLE 0–1*, and *NLE 0–2* is 185 students (I = 86, C = 99). Fixed effects and variances shown as parameter estimates with standard errors in parentheses. The models nested only intervention students within groups; comparison students were unclustered. ICCs estimated only for intervention students nested within instructional groups. The degrees of freedom (*df*) for tests of condition effects were based on the Satterthwaite approximation. Likelihood ratio tests, at bottom, compared homoscedastic residuals to heteroscedastic residuals with a criterion α of .20 and one degree of freedom.

****Significant at *p* = .0001; ***significant at *p* = .001; **significant at *p* = .01; *significant at *p* = .05.

Table 7

Student Posttest Adjusted Means

	Intervention (*n* = 87)			Comparison (*n* = 99)		
Posttest Measures	Unadjusted Mean	Adjusted Mean	Unadjusted SD	Unadjusted Mean	Adjusted Mean	Unadjusted SD
TUF-4	16.84	16.74	4.16	13.46	13.49	4.13
TUF-5	8.36	8.28	3.92	5.86	5.92	3.25
Test of Fraction Procedures	26.42	26.33	10.79	15.40	15.54	9.46
Curriculum-aligned Measure	16.16	16.13	4.66	11.77	11.80	3.55
NLE 0–1	90.65	90.76	7.53	80.03	79.97	11.67
NLE 0–2	87.24	87.31	8.33	80.82	80.76	8.15

Note. Total sample size *N* = 186 students (I = 87, C = 99). Sample size for *TUF-5*, *Test of Fraction Procedures (Full)*, *NLE 0–1*, and *NLE 0–2* posttests is 185 students (I = 86, C = 99).

Sensitivity analyses. We used a partially nested model with intervention students nested within tutoring groups and unclustered comparison students as the main model for impact analysis. This test of condition did not involve any further nesting levels (blocking), such as teacher or school. As a sensitivity analysis, we ran two other models: (a) students nested within tutoring groups and unclustered comparison students blocked by teacher, and (b) students nested within tutoring groups and unclustered comparison students blocked by school. We also tested for students nested within tutors (instead of tutoring groups) and unclustered comparison students (no blocks). In addition, as the *TUF-4* was given as both a pre- and posttest, we also examined the impacts from a gain score analysis for just the *TUF-4* outcome. Finally, we tested the model without covariates. All these tests demonstrate the sensitivity of condition effects to the various methodological variations. As seen from Table 8, all the various analyses resulted in similar impacts. We found no meaningful differences in the results due to the differences in the analytical approach. All *p*-values and effect sizes were similar across models. Moreover, the models that blocked on teachers and schools did not suggest any treatment-effect variation across the higher levels.

Table 8

Sensitivity of Main Effects to Analytic Approaches with Different Assumptions

Analysis Approach[a]	Analysis Unit	Block	Coefficient	(*SE*)	*p*-value	Hedges' *g*
			TUF-4			
ANCOVA	Small group[b]	–	3.25	(0.56)	< .0001	0.784
ANCOVA	Small group	Teacher	3.47	(0.51)	< .0001	0.84
ANCOVA	Small group	School	3.26	(0.59)	< .0001	0.79
ANCOVA	Tutor	–	3.45	(0.61)	< .0001	0.83
Gain Score[c]	Small group	–	3.25	(0.73)	< .0001	0.79
ANOVA	Small Group	–	3.29	(0.72)	< .0001	0.79
			TUF-5			
ANCOVA	Small group[b]	–	2.36	(0.46)	< .0001	0.660
ANCOVA	Small group	Teacher	2.36	(0.48)	< .0001	0.66
ANCOVA	Small group	School	2.29	(0.44)	.0001	0.64
ANCOVA	Tutor	–	2.29	(0.53)	.0002	0.64
ANOVA	Small Group	–	2.48	(0.64)	.0004	0.69
			Test of Fraction Procedures			
ANCOVA	Small group[b]	–	10.79	(1.37)	< .0001	1.068
ANCOVA	Small group	Teacher	10.89	(1.15)	< .0001	1.08
ANCOVA	Small group	School	10.99	(1.33)	< .0001	1.09
ANCOVA	Tutor	–	10.58	(1.69)	< .0001	1.05
ANOVA	Small Group	–	11.02	(1.79)	< .0001	1.09
			Curriculum-aligned Measure			
ANCOVA	Small group[b]	–	4.33	(0.51)	< .0001	1.055
ANCOVA	Small group	Teacher	4.45	(0.48)	< .0001	1.08
ANCOVA	Small group	School	4.39	(0.55)	< .0001	1.07
ANCOVA	Tutor	–	4.34	(0.59)	< .0001	1.06
ANOVA	Small group	–	4.37	(0.70)	< .0001	1.07

Analysis Approach[a]	Analysis Unit	Block	Coefficient	(*SE*)	*p*-value	Hedges' *g*
			NLE 0–1			
ANCOVA	Small group[b]	–	10.79	(1.20)	< .0001	1.083
ANCOVA	Small group	Teacher	10.59	(1.27)	< .0001	1.06
ANCOVA	Small group	School	10.70	(1.25)	< .0001	1.07
ANCOVA	Tutor	–	10.82	(1.17)	< .0001	1.09
Gain Score[c]	Small group	–	12.51	(1.66)	< .0001	1.26
ANOVA	Small group	–	10.64	(1.39)	< .0001	1.07
			NLE 0–2			
ANCOVA	Small group[b]	–	6.56	(0.95)	< .0001	0.796
ANCOVA	Small group	Teacher	6.57	(0.93)	< .0001	0.80
ANCOVA	Small group	School	6.56	(0.98)	< .0001	0.80
ANCOVA with tutor as cluster	Tutor	–	6.62	(0.75)	< .0001	0.80
ANOVA	Small group	–	6.38	(1.25)	< .0001	0.78

[a]All ANCOVA models included the *WRAT-4* and *NLE 0–1* variables as covariates. [b]Main model used in the impact analysis. [c]Time × Condition analysis.

Impact on performance assessments. Five performance-based assessments were administered throughout the course of the intervention to all participating intervention students and to a randomly selected sub-sample of 34 comparison students. To compare the relative performance of both student groups, we randomly selected 35 intervention students and compared their performance to that of the 34 comparison students. As seen in Table 9, intervention students significantly outperformed comparison students on all five performance assessments (g = .68 to 1.23), and p-values were all statistically significant after the Benjamini and Hochberg (1995) correction for multiple comparison.[3]

[3]Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological, 57*(1), 289–300.

Table 9

Impact on Five Performance Assessments

Performance Assessment	Possible points	Intervention (*n* = 35) Mean raw score	SD	Comparison (*n* = 34) Mean raw score	SD	*t*-test *(df)*	*p*-value	Hedges' *g*
1. Estimating a Fraction on a Number Line								
Composite Score	5	2.82	1.47	1.68	1.41	3.29 (66)	.002	.789**
Item 1: Correct Answer	2	1.62	0.74	1.21	0.91			
Item 2: Explanation	3	1.21	0.98	0.47	0.79			
Number of Math Vocabulary Words used	--	1.21	1.15	0.35	0.49			
2. Ordering Fractions on a Number Line								
Composite Score	6	1.60	1.77	0.53	1.28	2.87 (67)	.006	.683**
Item 1: Correct Answer	2	0.60	0.85	0.24	0.61			
Item 2: Explanation	4	1.00	1.21	0.29	0.80			
Number of Math Vocabulary Words used	--	2.37	1.44	0.62	0.74			
3. Fraction Addition Word Problem								
Composite Score	9	3.60	1.58	2.03	1.57	4.15 (67)	.000	.988***
Item 1: Correct Answer	3	0.74	0.56	0.26	0.71			
Item 2: Explanation	5	1.89	1.13	0.85	0.93			
Item 3: Correct Operation	1	0.97	0.17	0.91	0.29			
Number of Math Vocabulary Words used	--	1.86	1.09	0.47	0.66			
4. Fraction Multiplication Word Problem								
Composite Score[a]	6	2.20	2.08	0.24	0.78	5.21 (44.2)	.000	1.228***
Item 1: Correct Answer	1	0.57	0.50	0.06	0.24			
Item 2: Explanation	4	1.09	1.17	0.09	0.29			
Item 3: Correct Operation	1	0.54	0.51	0.09	0.29			
Number of Math Vocabulary Words used	--	2.17	1.15	0.59	0.78			
5. Fraction Subtraction Word Problem								
Composite Score[a]	8	3.24	2.20	1.38	1.30	4.22 (54.8)	.000	1.012***
Item 1: Correct Answer	2	0.56	0.82	0.09	0.38			
Item 2: Explanation	5	1.97	1.34	0.71	0.84			
Item 3: Correct Operation	1	0.71	0.46	0.59	0.50			
Number of Math Vocabulary Words used	--	2.50	1.46	0.97	1.22			

Note. Total analytic sample = 69 students (I = 35, C = 34). Sample size for Performance Assessments 1 and 5 is 68 students (I = 34, C = 34). [a]Unequal variances; used two sample *t*-tests with Welch's Approximation and corresponding Hedges' *g*. ***Significant at $p = .001$; **significant at $p = .01$.

***Quality of explanations*.** Across all five performances, the explanations given by intervention students were more thorough than the explanations provided by comparison students. Intervention students' written explanations included more math vocabulary words, more coherent rationales for their decisions and solutions methods, and details that demonstrated their understanding of the procedures they employed. For example, intervention students used an average of 2.02 relevant math vocabulary words, such as *common denominator*, *simplify*, and *unit fraction*, compared to an average of 0.6 words used by comparison students. The following is typical of the explanations that were provided by comparison students for the fifth performance assessment (*Manny is training for a marathon. He ran* $10\frac{2}{5}$ *miles on Saturday. On Sunday, he ran* $7\frac{9}{10}$ *miles. How many more miles did Manny run on Saturday than Sunday?*): "I just multiplyed the two number[s] together. Then I made sure that I did it right. And that is how I got my answer." In contrast, the explanations provided by intervention students were more detailed, organized, and comprehensive (the underlined words are mathematical vocabulary words emphasized in the *TransMath* curriculum):

> For my answer I got $2\frac{5}{10}$ then I <u>simplified</u> it to $2\frac{1}{2}$ so my answer was a <u>mixed number</u> and $2\frac{1}{2}$. The first thing I did was look for a clue word and that was how many more, so I knew I had to subtract. Next, I lined up my <u>fractions</u> and <u>whole numbers</u> correctly. Then, I checked if they had a <u>common denominator</u>, which they didn't. Then I noticed 10 was a multiple of 5 and kept $7\frac{9}{10}$ then changed $10\frac{2}{5}$ to $10\frac{4}{10}$. Finally, I subtracted and saw that you couldn't subtract $\frac{4}{10}$ minus $\frac{9}{10}$, so I regrouped. I regrouped by taking a whole away from ten and added $\frac{10}{10}+\frac{4}{10}$ to get

$\frac{14}{10}$ and change 10 into 9. Then I got the answer of $2\frac{5}{10}$ which is equivalent to $2\frac{1}{2}$.

That's how I got my answer.

See Appendix A for additional examples of student written explanations.

4.2 *CLASS* Observations of Tutoring Groups

To better understand the facets of the learning environment that led to enhanced outcomes for this at-risk group of students, observations of the tutoring groups were conducted using the *CLASS* observational system (Pianta et al., 2012a). Instruction during the 35-minute *TransMath* fractions intervention lesson was rated on a 7-point Likert scale, where a rating of 1 or 2 is low range; 3, 4, 5 encompasses the middle range; and 6, 7 indicates a high performance level. Mean rating scores and the range of scores across the 21 tutoring groups are presented in Table 10. Mean ratings were in the high range only for mean Behavior Management and Productivity and in the high-middle range for Teacher Sensitivity, Instructional Formats, and Student Engagement.

Interpreting ratings of tutoring groups. Table 10 also includes *CLASS* ratings from the *Measures of Effective Teaching Study (MET)*, which included a nationwide sample of 1,333 teachers from grades 4–6; these ratings were obtained from the Upper Elementary and Secondary *CLASS* Technical Manual study (Pianta, Hamre, & Mintz, 2012b). Note that the ratings from the *MET* study are not nationally representative but provide an example of scores from a large nationwide sample (i.e., a reference point), to facilitate comparison and an understanding of the ratings from this study. Also note that while the *MET* study observed videos of general education teachers frequently teaching whole classes of students, in the current study, tutors were observed working with small groups of approximately five students.

Tutors received a higher rating than the *MET* teachers for Student Engagement and the domain of Classroom Organization. They also scored higher in the Emotional Support domain for Positive Climate and Teacher Sensitivity; in the Classroom Organization domain for Behavior Management, Productivity, and Negative Climate; and in the Instructional Support domain for Instructional Learning Formats, Content Understanding, and Quality of Feedback.

Tutors scored lower than the *MET* teachers on Regard for Student Perspectives, Analysis and Inquiry, and Instructional Dialogue, which is not surprising, given the context and goals of the small-group tutoring setting. Time during the 35-minute tutoring session was spent on four structured activities: review, explicit instruction, guided practice, and student explanations. Lessons progressed at a rapid pace, with tutors explicitly teaching the fractions content. Time was allocated for eliciting student explanations, but in general, the sessions were not designed for in-depth, rich content-focused discussions and back and forth extended exchanges between tutors and students (i.e., Instructional Dialogues). There were brief but meaningful structured peer interactions, but the sessions were also not planned to provide opportunities for student autonomy and leadership (i.e., Student Perspectives). Also, while the fractions intervention did emphasize problem solving, it did not provide for extensive student explorations of novel or open-ended problems, tasks, or questions, or for other types of higher-order thinking activities (i.e., Analysis and Inquiry). These types of activities typically occur during core mathematics instruction—when they do occur. Students in both conditions did receive core mathematics instruction.

Table 10

CLASS Observation Ratings

	CLASS Rating: current study			*CLASS* Rating: *MET* study[a]		
CLASS Item	Mean	SD	Range	Mean	SD	Range
Emotional Support	4.38	0.97	2.0–6.3	n/a	n/a	n/a
Positive Climate	4.66	1.38	2–7	4.68	0.61	2.38–6.50
Teacher Sensitivity	5.72	1.16	2–7	4.26	0.55	2.33–6.50
Regard for Student Perspectives	2.75	0.92	2–5	3.29	0.60	1.38–5.18
Classroom Organization	6.77	0.38	5.7–7.0	n/a	n/a	n/a
Behavior Management	6.71	0.57	5–7	6.01	0.58	2.88–7.00
Productivity	6.72	0.56	5–7	5.91	0.46	3.88–7.00
Negative Climate	1.11	0.32	1–2	1.32	0.35	1.00–3.38
Instructional Support	4.00	0.72	2.2–5.2	n/a	n/a	n/a
Instructional Learning Formats	5.21	0.85	3–6	4.36	0.52	2.50–6.50
Content Understanding	4.72	0.86	3–6	3.97	0.53	2.08–6.50
Analysis and Inquiry	2.68	0.72	2–4	2.80	0.50	1.63–4.31
Quality of Feedback	4.21	1.07	2–6	3.76	0.57	2.00–5.81
Instructional Dialogue	3.17	0.78	1–4	3.51	0.56	2.00–5.50
Student Engagement	5.74	1.18	3–7	5.08	0.48	2.94–6.50

Note. Each of the 12 *CLASS* dimensions are rated on a 7-point Likert scale, with 1–2 indicating a low score, 3–5 being moderate, and 6–7 indicating a high score. Scoring for Negative Climate (Classroom Organization) is reversed; when calculating the domain total, it is necessary to subtract the average score for Negative Climate from 8.

[a]Data obtained from the Upper Elementary and Secondary *CLASS* Technical Manual for (Pianta et al., 2012b). The *Measures of Effective Teaching Study (MET)* was conducted nationwide with a sample of 1,333 teachers in Grades 4–6. Each teacher was assessed using *CLASS* throughout one academic year four to eight times. n/a = not available.

Correlations between *CLASS* ratings and student fractions outcomes.

Correlations between *CLASS* ratings and student fractions outcomes for students in the intervention group are presented in Table 11. Three dimensions—Quality of Feedback, Instructional Dialogue, and Content Understanding—were significantly correlated with all measures of fraction achievement: *TUF-4*, *TUF-5*, *Test of Fraction Procedures*, and the *Curriculum-aligned Measure*. Interestingly, none of the *CLASS* dimensions were correlated with the *NLE* measure.

Table 11

CLASS Correlations with Student Fractions Outcomes

CLASS Item	*TUF-4 Post*	*TUF-5 Post*	*Test of Fraction Procedures Post*	*Curriculum-aligned Post*	*NLE 0–1 Post*	*NLE 0–2 Post*
Emotional Support	−.035	.050	.043	.048	−.017	−.117
Positive Climate	−.040	.058	.029	.056	.047	−.087
Teacher Sensitivity	−.116	.005	−.050	−.024	−.068	−.161
Regard for Student Perspectives	.096	.063	.158	.099	−.040	−.036
Classroom Organization	.116	.211	.104	.152	−.008	−.015
Behavior Management	.138	.210	.098	.079	.041	.022
Productivity	.040	.131	.085	.176	−.068	.002
Negative Climate	−.091	−.136	−.041	−.083	.017	.093
Instructional Support	.194	**.232***	**.233***	**.240***	.026	.020
Instructional Learning Formats	−.023	.102	.042	.115	−.072	−.106
Content Understanding	.147	.201	**.237***	**.276****	.025	.041
Analysis and Inquiry	.187	.124	.115	.116	.120	.057
Quality of Feedback	**.253***	**.265****	**.260***	**.217***	.047	.056
Instructional Dialogue	**.238***	**.254***	**.299****	**.274****	−.006	.037
Student Engagement	−.018	.071	.001	.065	−.060	−.128

Note. Total sample size N = 87 intervention students. Sample size for *TUF-5*, *Test of Fraction Procedures*, *NLE 0–1*, and *NLE 0–2* is 86 intervention students. Each of the 12 *CLASS* dimensions are rated on a 7-point Likert scale, with 1–2 indicating a low score, 3–5 being moderate, and 6–7 indicating a high score. Scoring for Negative Climate (Classroom Organization) is reversed. Computed correlation used Pearson method with pairwise deletion. *Significant at p = .05; **significant at p = .01.

Quality of Feedback addresses the extent to which tutor feedback increases and extends students' understanding and learning and encourages students to participate. Ratings are based on behavior markers that reflect the extent to which tutor feedback enhances learning. The behavior markers include feedback loops (e.g., back-and-forth exchanges between tutor and student), scaffolding (e.g., hints and prompting), building on student responses (e.g., specific feedback, clarification), and encouragement and affirmation (e.g., recognizing and affirming student effort).

Instructional Dialogue reflects the practice of purposeful discussions between tutor and students. The primary purpose of this dimension is to determine if connections between and among ideas are made to enhance students' understanding of the content at hand. Behavior markers include cumulative content-driven exchanges (e.g., depth of the exchanges, exchanges that build on one another rather than switch topics), distributed talk (e.g., a balance between tutor and student talk, majority of students have the opportunity to contribute), and facilitation strategies (e.g., use of open-ended questions and statements; acknowledging, repeating, and extending student responses).

The middle to low-middle mean ratings for these two dimensions of instruction (Quality of Feedback = 4.21; Instructional Dialogue = 3.17) indicate that a modest amount of feedback and dialogue occurred during the intervention lessons. Most importantly, these levels of elegant feedback and instructional dialogue were significantly correlated with several of the student mathematics outcomes (with correlations ranging from .217 to .299; $p < .05$).

These statistically significant associations were, in all likelihood, stimulated by the instructional design of the fractions intervention lessons. Specifically, throughout the lessons students are asked to explain the reasoning behind their answers, which lends itself to the type of feedback, prompting, and instructional dialogue measured by *CLASS* and recommended by contemporary state standards and the CCSS Practice Standards.

For example, after explaining and modeling ways to add fractions with like denominators, the tutor directed students to write $\frac{3}{4} + \frac{3}{4} = \frac{6}{8}$ on their whiteboard. Next, students were asked if the problem was solved correctly. Together, the students and tutor discussed how the equation could be "fixed." After students worked in pairs to solve the problem, another discussion took place after students were asked, "Why is that correct and the first way incorrect?" Students discussed their responses using either Cuisenaire Rods or a number line to illustrate their rationale. Clearly, the instructional dialogue and feedback generated by this type of interactive instruction was beneficial to the students and was significantly positively correlated with student performance on the fractions outcome measures.

The average rating of 4.72 on the Content Understanding borders on the upper end of the middle range. Content Understanding was significantly correlated with one fractions posttest: the *Test of Fraction Procedures* (r = .237; p < .05). Content Understanding dimension captures the complexity of the lesson's content and the instructional methods used to help students understand key attributes of the concepts. It focuses on building student understanding of the relationships among facts, skills, and concepts; establishing real-world connections; identifying essential components and communicating them through multiple examples and contrasting examples; linking to prior knowledge; and attending to misconceptions.

For example, when students were initially taught to add fractions with unlike denominators, the concept was related to a previously learned skill, solving problems with like denominators. Thus, tutors addressed behavior markers such as tapping prior knowledge, providing contrasting examples, multiple and varied examples, and cumulative review, all of which assist in decreasing student misconceptions. The tutors provided explicit instruction in using the algorithm for solving addition problems containing fractions with unlike denominators. The instruction also included an explanation of why the algorithm works using visual representations.

Classroom organization and student engagement. Observational research in general has shown both Classroom Organization and Student Engagement to be correlated with student outcomes during whole-class instruction (Foorman et al., 2006;

Gersten et al., 2009; Jayanthi et al., 2017). However, in this study, despite their high ratings, Classroom Organization and Student Engagement were not significantly correlated with student fraction outcomes.

Classroom Organization received an exceptionally high mean rating of 6.77. During the 35-minute session, tutors managed time and instructional routines so that time allocated to instruction was maximized. It also indicates that the tutor set up clear expectations and managed student behavior effectively, and that there were very few incidences of unwanted negative behavior. Of course, this rating is easier to achieve in a small group of five or so than in a full class. Nonetheless, this finding suggests that the interventionists provided an organized learning environment for their students.

The rating for Student Engagement was in a high-middle range (5.74), indicating that the majority of the students were engaged or on-task. However, there was variability in this dimension. Observation data indicated that some students were actively attending (e.g., answering questions, sharing ideas) and participating in the instructional activity, some were engaging passively (e.g., watching/listening to the tutor), and others were disengaged for part of the time. Although one would assume that student engagement would be easier to achieve in a small group of five than in a class of 30, that has not always been the case (e.g., Gersten, Carnine & Williams, 1982).

4.3 Exploratory Analyses

Moderator analysis. To determine if there were any differential impacts, student pretest scores (*NLE 0–1*, *TUF-4*, *WRAT-4*, *Test of Fraction Procedures [Add/Sub]*), demographic variables (gender, free and reduced lunch), and districts (Districts 1, 2, and 3) were examined as moderators of the relation between the fractions intervention and student achievement at posttest. There were no statistically significant tests of moderation for gender.

We found nine statistically significant moderators. As the moderator analyses are correlational and exploratory in nature, and not causal and confirmatory, we did not correct for multiple tests.

Student pretest moderators. Treatment students outperformed comparison students on all outcome measures. Furthermore, student scores on the *WRAT-4* pretest significantly moderated intervention effects on the *Curriculum-aligned Measure* scores (p = .0075), *TUF-5* posttest scores (p = .0264), and the *NLE 0–1* posttest scores (p = .0132). Results indicated that higher *WRAT-4* pretest scores were associated with a larger difference between treatment and comparison students' scores on the *Curriculum-aligned Measure* and *TUF-5* posttest. In contrast, lower *WRAT-4* pretest scores were associated with a larger difference between treatment and comparison students' scores on the *NLE 0–1* posttest.

Student scores on the *NLE 0–1* pretest moderated the intervention effects on the *Curriculum-aligned Measure* scores (p = .0087) and the *NLE 0–1* posttest scores (p = .0204). Higher performance on the *NLE 0–1* pretest measure was associated with a larger difference between treatment and comparison students' scores on the *Curriculum-aligned Measure*, while lower performance on the *NLE 0–1* pretest measure was associated with a larger difference between treatment and comparison students' scores on the *NLE 0–1* posttest.

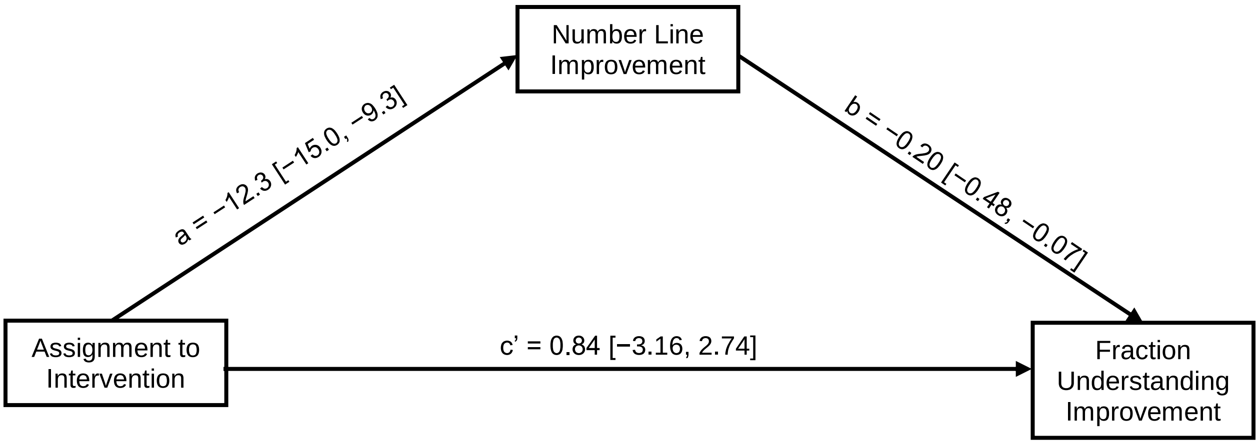
Demographic moderators. Free and reduced-price lunch status also significantly moderated intervention effects on the *TUF-4* (p = .0374) and *NLE 0–1* (p = .0400) posttests. Although treatment students outperformed comparison students on both the *TUF-4* pretest and the *NLE 0–1* pretest regardless of free and reduced-priced lunch status, the difference between treatment and comparison students' scores on the *TUF-4* posttest was more pronounced among students who were not receiving free and reduced-price lunch, compared with students who were receiving free or reduced-priced lunch. The opposite was true for *NLE 0–1* post-test performance—there was a larger difference between treatment and comparison students' scores on the *NLE 0–1* posttest among students who were receiving free and reduced-price lunch, compared with students who were not receiving free and reduced-price lunch.

School district moderators. Students in District 2 performed lower than students from other districts on all the pretest measures, including the *TUF-4* screening measure (p = .0048, .0000, .0774, and .0032 for the *TUF-4*, *WRAT-4*, *Test of Fraction Procedures [Add/Sub]*, and *NLE 0–1* pretests, respectively). Meanwhile, students in District 3 performed higher than students from Districts 1 and 2 on all pretest measures (p = .0000, .0012, .0463, .0000 for the *TUF-4*, *WRAT-4*, *Test of Fraction Procedures [Add/Sub]*, and *NLE 0–1* pretests, respectively). Although treatment students from District 2 began the study at a lower achievement level than the other districts, their level of understanding grew at a faster rate than predicted over the course of the study. School district membership significantly moderated the size of the effect of the intervention on *NLE 0–1* posttest scores. The difference between treatment and comparison students' *NLE 0–1* posttest scores was largest among students from District 2 (District 2 vs. 1 and 3; p = .0181). In contrast, the gap between treatment and comparison students' *NLE 0–1* posttest scores was smallest among students from District 3 (District 3 vs. 1 and 2; p = .0213). This finding suggests that treatment students from the lowest-performing district seemed to benefit the most from the intervention. This finding could be related to the degree to which number lines were used during core mathematics instruction. Teachers from District 3 reported that core classroom instruction incorporated the use of number lines to teach fraction content (i.e., magnitude, equivalence, comparing, and the four operations) more often than did teachers from District 2. Of the seven teachers from District 3, 57% reported using number lines to teach fraction magnitude, 100% reported using number lines to teach fraction comparing, and 71% reported using number lines to teach fraction addition and subtraction, compared to 45%, 73%, and 55% of the 11 teachers from District 2. Because students from District 2 had limited exposure to using number lines to understand fractions concepts during core mathematics instruction, the use of these materials during the intervention could have acted as a springboard, boosting these treatment students to much higher achievement levels than predicted. In contrast, students from District 3 were familiar with using number lines outside of the intervention, so the difference between treatment and comparison students' scores was less pronounced.

Mediator analysis. We tested whether the effects of condition on gains on *TUF-4* were potentially mediated by gains on *NLE 0–1*. Figure 1 depicts an indirect-path model, the results of which offer evidence that supports mediation. The direct effect of condition on gains in *TUF-4* were statistically significant, 3.34 [2.13, 4.55], an estimate very similar to the results in Tables 6, 7, and 8. With the indirect path a·b in the model, the effect of condition on gains in *TUF-4* dropped to 0.84 and was no longer statistically significant (confidence bounds exclude zero). Thus, the results support the fully mediated effect of intervention versus comparison on students' fraction understanding (i.e., *TUF-4* gains) through improvement on the number line measure. Mediation models are, however, correlational and cannot support a directional or causal interpretation.

Figure 1

Mediation Path Model



This mediation model demonstrates that after accounting for the indirect effects of condition on gains in fraction understanding through gains on the number line measure, the direct effects of condition on gains in fractions understanding become statistically nonsignificant. The model, therefore, supports the hypothesis of complete mediation by number line improvement. The paths provide the raw estimates and 95% confidence intervals.

Pre- posttest correlations. Correlations between pre- and posttests are presented in Table 12. The *WRAT-4*, a measure of general mathematics achievement, was moderately and significantly correlated with all the posttests. The correlations ranged from .474 to .568, an optimal range for a covariate (Keppel & Zedeck, 1989). The *NLE 0–1* pretest was also moderately and significantly correlated with all the posttests, with correlations ranging from .393 to .534. *TUF-4* did not correlate as well with the posttests (r = .140 to .264), perhaps because it was used as a screener and therefore the scores of the students included in the study had a severely restricted range.

Table 12

Correlations Between Pre- and Posttest Student Measures

	TUF-4 Post	*TUF-5 Post*	*Test of Fraction Procedures Post*	*Curriculum-aligned Post*	*NLE 0–1 Post*	*NLE 0–2 Post*
TUF-4 Pre	.190**	.140	.204**	.264***	.211**	.161*
WRAT-4 Pre	.568***	.560***	.561***	.493***	.474***	.505***
Procedures (Add/Sub) Pre	.307***	.260***	.332***	.258***	.223**	.290***
NLE 0–1 Pre	.507***	.429***	.446***	.393***	.477***	.534***

Note. Total sample size N = 186 students (I = 87, C = 99). Sample size for *TUF-5*, *Test of Fraction Procedures*, *NLE 0–1*, and *NLE 0–2* is 185 students (I = 86, C = 99). Computed correlation used Pearson method with pairwise deletion.
***Significant at p = .001; **significant at p = .01; *significant at p = .05.

4.4 Understanding Implementation

Intended versus actual implementation of the fractions intervention. The 52 *TransMath* fractions lessons were intended to be 35 minutes each and delivered 3 to 4 times per week in small groups of 5 students. In actuality, the overall average session length was 34 minutes (median = 34:00 minutes; range = 14 to 54 minutes). Although every participating teacher originally consented to the 35-minute tutoring block, several teachers later insisted on restricting the lessons to a block of 25 to 30 minutes. This was mainly an issue at District 1 and at two schools in District 2.

The actual mean number of sessions per week ranged from 2.5 to 3.1 lessons across the 21 tutoring groups (median ranged from 2 to 4 lessons per week). Most of the tutoring groups began the intervention on a 4-day schedule; however, several classroom teachers requested a 3-day tutoring schedule after the first few weeks of the intervention to accommodate unexpected schedule changes within their math departments. Tutoring lessons frequently had to be rescheduled, most frequently at District 2 due to an abundance of unanticipated field trips and school assemblies. In addition, the designated tutoring rooms were often unavailable at this district, and the tutoring sessions had to be rescheduled. Unexpected school activities and room changes were not an issue in the schools in District 3, which had designated tutoring rooms available and intervention time blocks incorporated into the weekly school schedule.

At randomization, each of the 21 tutoring groups consisted of 4 to 5 students; however, after attrition, group sizes ranged from 2 to 5 students. There was only one case in which a tutoring group had only 2 students. The median group size was 4 students (mean = 4.24; *SD* = 0.89).

Fidelity of implementation. For each of the 21 tutoring groups, fidelity was assessed for eight lessons (Lessons 4, 11, 20, 27, 31, 35, 43, and 47) by a member of the research team. Findings from both procedural fidelity as well as quality of instruction are described next.

***Procedural fidelity*.** Across the 21 tutoring groups (and across all eight lessons), on average, 78.05% of the steps were completed by the tutors (median = 81.08%; range = 69.57–85.07%). Of the 21 groups, eight groups completed over 80% of the steps (range = 80.14% to 92.91%; median = 88.08%), and 10 groups completed between 70 and 80% of the steps (range = 70.47 to 79.19; median = 79.31%). Fidelity for three groups was below 70% (range = 51.72% to 67.59%; median = 60.29%).

When examined by lesson, the highest mean fidelity was recorded for Lesson 4: *Representing Fractions with Cuisenaire Rods*, with 85.07% of the steps completed. The lesson with the lowest mean procedural fidelity was Lesson 31: *Multiplying Proper Fractions Using an Area Model* (mean = 69.57%; range = 51.22–85.37%). Thirteen groups' implementations of Lesson 31 had procedural fidelity less than or equal to 75.00%. Four tutoring groups had procedural fidelity less than or equal to 60.00%. One reason for this low fidelity for Lesson 31 might be related to the number of activities that had to be completed during this session. It was a complex lesson that focused on the key concept of fraction multiplication. It was also a dense lesson, with many key concepts and activities that had to be covered. Compared to the 37 steps that had to be implemented in Lesson 4, 82 steps had to be completed in Lesson 31. Perhaps it was a combination of these issues—difficult topic combined with a content-intense lesson—that made it difficult for the tutors to teach all the aspects of the lesson in the given time.

***Quality of instruction during implementation*.** Fidelity was also assessed by examining the quality of instruction provided in each tutoring group. Six items were rated on a 5-point Likert scale, with 1 being Low and 5 being High. Mean quality ratings for each instructional behavior are presented in Table 13. Overall, across the 21 groups and the eight fidelity lessons, the average quality rating was 3.99. Nine groups had a rating greater than 4 (median = 4.50). The rating for 11 groups was between 3 and 4 (median = 4.00). Only one group received a rating of 2.86, indicating that in most of the groups, the quality of instruction provided by the tutors was above average.

Overall quality ratings were lowest for Lesson 43: *Converting Mixed Numbers to Fractions Over One* (mean = 3.83; median = 4). For this lesson, groups received ratings as low as 1 and 2 on five of the six quality attributes. Lesson 20: *Fractions That Represent the Same Number* had the highest overall quality rating (mean = 4.12; median = 4.5). For this lesson, eight of the 21 tutoring groups received ratings above 4 on all six quality attributes, and 12 of the 21 groups received ratings above 4 on at least five of the six quality attributes.

Across all the groups and fidelity lessons, the behavior that received the lowest quality rating (a rating of 1 or 2) most often was providing specific math-oriented praise (mean = 3.34; median = 3). The two instructional behaviors with the highest mean quality ratings were maintaining a positive rapport with the students (mean = 4.52) and using clear and mathematically correct language (mean = 4.34).

As probing for understanding and facilitating student discussion were an important aspect of the intervention, the fidelity observers also rated each group on whether the tutor asked open-ended questions to probe student thinking and understanding before providing an answer or solution. The fidelity observers indicated that, across all tutoring groups and lessons, tutors probed for student understanding *frequently* 64% of the time and *sometimes* 30% of the time.

Table 13

Quality of Implementation Across All Eight Lessons for All 21 Tutoring Groups

Fidelity Item	Mean	SD	Range	Median
Observer's overall rating of the tutor's implementation.	3.99	0.83	3.83–4.12	4.00
Observer's perception of students' grasp of the content.	4.08	0.85	3.98–4.17	4.00
Tutor paces the lesson so that all parts of the session were covered in sufficient depth.	3.84	1.03	3.71–4.17	4.00
Tutor uses clear and mathematically correct language.	4.34	0.74	4.14–4.62	4.50
Tutor enhances students' explanations.	4.13	0.90	3.90–4.31	4.00
Tutor provides specific math-oriented praise.	3.34	1.34	2.76–4.17	3.00
Tutor Maintains a positive rapport with students.	4.52	0.75	4.36–4.76	5.00

	% Rarely	% Sometimes	% Frequently
When students are explaining their answers, the tutor asks open-ended questions to probe thinking and understanding before providing an answer or solution.	6.02	30.12	63.86

Note. Each of the quality of implementation items are rated on a 5-point Likert scale, with 1 indicating a low score, 3 being moderate, and 5 indicating a high score.

Reliability of the fidelity data. Inter-rater reliability on fidelity scoring was assessed on 13.10% of the sessions (n = 22), by having two raters independently assess the fidelity. The mean inter-rater reliability for the procedural fidelity was 81.57% (median = 81.03%). The mean inter-rater reliability for quality of implementation ratings was 75.97% (median = 85.71%).

Correlation between fidelity and student fractions outcomes**.** Both procedural fidelity and quality of implementation ratings were moderately and significantly correlated with *TUF-4*, *TUF-5*, and the *Test of Fraction Procedures* (r = .372 to .450, p < .001). See Table 14. Fidelity was correlated to a lesser extent, but still significantly at p < .01, with the *Curriculum-aligned Measure* (r = .324 and .271 for procedural fidelity and quality of implementation, respectively). While correlations between fidelity of implementation and *NLE 0–1* were very low and non-significant, correlations with *NLE 0–2* were low but significant at p < .05 (r = .242 and .246 for procedural and quality fidelity of implementation, respectively).

Table 14

Correlation Between Fidelity of Implementation and Student Fraction Outcomes

Fidelity	*TUF-4 Post*	*TUF-5 Post*	*Test of Fraction Procedures Post*	*Curriculum-aligned Post*	*NLE 0–1 Post*	*NLE 0–2 Post*
Procedural	.379***	.449***	.450***	.324**	.043	.242*
Quality of Implementation	.372***	.423***	.414***	.271**	.114	.246*

Note. Total sample size N = 87 intervention students. Sample size for *TUF-5*, *Test of Fraction Procedures*, *NLE 0–1*, and *NLE 0–2* is 86 intervention students.
***Significant at p = .001; **significant at p = .01; *significant at p = .05.

4.5 Findings from Surveys

Classroom teacher survey of instructional practices during core mathematics instruction. Core classroom mathematics teachers varied in terms of when they started teaching fractions to their students. Variations were seen within and across districts. At District 1, 63.64% of teachers began teaching fractions in spring, and the rest started in fall and winter (September–December 2016 = 9.09%; January–February 2017 = 27.27%). In District 2, fraction instruction started at different time points in fall, with all teachers having begun by January 2017 (September 2016 = 18.18%; November 2016 = 45.45%; December 2016 = 9.09%; January 2017 = 27.27%). In District 3, a majority of the teachers (85.71%) started teaching fractions in November and December 2016, and the rest (14.29%) started in January 2017.

My Math (McGraw-Hill Education, 2017b) was used as the core curriculum in District 1, *California Math* (McGraw-Hill Education, 2017a) was used in District 2, and *GO Math!* (Houghton Mifflin Company, 2018) was used in District 3. Ninety percent of the teachers indicated that they used a district-adopted textbook; 83 percent used supplemental material including lessons/problems developed by the district (31.03%).

To facilitate a comparison between the fractions intervention and the core classroom fractions instruction, teachers were asked several specific questions regarding the fractions topics they teach, the fractions operations they focus on, the representations and methods they use, and their expectations for their students regarding explanations and discussions. Responses to these questions are summarized in Table 15. Note that items that are starred in the table are also addressed in the *TransMath* fractions intervention provided to the students.

Table 15

Classroom Teacher Survey of Core Mathematics Instruction

Survey Item	Percentage of Teachers
I use the following representations to teach fractions:	
Concrete manipulatives*	75.86
Visual representations*	96.55
Number lines*	96.55
I use number lines to teach the following fraction content:	
Fraction magnitude*	51.72
Fraction equivalence*	96.55
Fraction comparing*	86.21
Fraction addition*	65.52
Fraction subtraction*	65.52
Fraction multiplication*	31.03
Fraction division*	31.03
I teach the following fraction operations:	
Add/subtract fractions or mixed numbers with unlike denominators*	96.55
Multiply a fraction times a fraction*	96.55
Divide a fraction by a fraction*	93.10
Add/subtract mixed numbers that require whole-number regrouping*	89.66
Multiply mixed numbers*	89.66
I use the following methods for teaching students how to compare or order fractions when evaluating magnitude:	
Cross multiplying	79.31
Number line*	93.10
Benchmark fractions*	79.31
Finding common denominators*	89.66
Drawing a picture of each fraction	89.66
Use of manipulatives*	82.76
Thinking about the meaning of the numerator (part) and the denominator (whole)*	89.66
Thinking about the relative size of the numerator compared to the denominator*	72.41
I use the following methods when teaching students how to solve fraction word problems:	
Draw a picture	100.00
Think about problem types	82.76
Make a table	65.52
Write an equation*	93.10
Focus on key words	100.00

	Never	*Rarely*	*Sometimes*	*Often*	*Always*
I require students to explain their thinking after solving a problem.*	0.00	3.45	24.14	48.28	24.14
I require students to explain using mathematically valid vocabulary.*	0.00	0.00	6.90	51.72	41.38
I use prompt cards to help guide students as they form explanations.*	55.17	17.24	20.69	3.45	3.45
Students explain their thinking or understanding of fractions concepts *orally*.*	0.00	3.45	17.24	58.62	17.24
Students explain their thinking or understanding of fractions concepts *in writing*.*	0.00	3.45	17.24	58.62	17.24
Students get to work on novel word problem-solving activities related to fractions.*	6.90	10.34	55.17	24.14	0.00

Note. Total Analytic Sample *N* = 29 teachers. Some percentages do not sum to 100 because some teachers selected more than one answer. Items that are starred in the table are also addressed in the *TransMath* fractions intervention provided to the students.

Almost all classroom mathematics teachers reported using visual representations including number lines (96.55%); about three-fourths indicated they also used concrete manipulatives. Responses to the question regarding the use of number lines to teach fraction content were of particular interest as the number line was a critical tool used throughout the *TransMath* intervention. All but one teacher reported using a number line, but the extent of use varied dramatically. While a majority of the teachers used the number line to teach fraction equivalence (96.55%) and fraction comparison (86.21%), only 51.72% used it to teach fraction magnitude. Teachers were also using the number line to teach fractions operations. About two-thirds were using it to teach addition and subtraction (65.52%), but a much smaller percentage used number lines for division and multiplication of fractions (31.03%).

Another important element of the fractions intervention was the emphasis on student explanations as a vehicle for determining student understanding. A majority of the core classroom mathematics teachers indicated that they required their students to explain their thinking after solving a problem, but the frequency varied (*always*: 24.14; *often*: 48.28%; *sometimes*: 24.14%). As in the fractions intervention, they also required their students to explain using mathematically valid vocabulary (*often*: 51.72; *always*: 41.38%). Teachers reported having students provide explanations orally as well as in writing.

Another key aspect of the *TransMath* fractions intervention is the focus on novel problem-solving activities. Over 50% of the classroom mathematics teachers indicated that their students worked on novel problems *some* of the time; another 24.14% indicated that they *often* worked on such problems.

One difference between the *TransMath* fractions intervention and core classroom mathematics instruction was the use of prompt cards to help guide students as they form their explanations. Most of the teachers reported *never* (55.17%) or *rarely* (17.24%) using the prompt cards.

***Instructional activities of comparison students*.** Classroom mathematics teachers were asked to provide a description of the activities their comparison students engaged in during the 35-minute intervention block when intervention students were receiving the fractions intervention. Students were engaged in a variety of instructional and intervention activities. See Table 16. As anticipated, very few teachers (17.24%) reported providing a structured mathematics intervention.

Table 16

Activities of Comparison Students During the 35-minute Intervention Block

What are the rest of your students doing while the selected students go to fractions tutoring?	Percentage of Teachers
Math intervention	17.24
Reading intervention	17.24
Core math	31.03
Other instruction (ELA, Science, Social Studies)	72.41
Other (P.E., homework review)	17.24

Note. Total Analytic Sample *N* = 29 teachers. Percentages do not sum to 100 because some teachers selected more than one answer.

Appraisal surveys. Findings from student, tutor, and classroom teacher appraisal surveys are presented below.

Student survey**.** Virtually all of the fifth-graders (96.55%) indicated that they liked attending the fractions tutoring group all the time or some of the time. Over 85% indicated that they found fractions tutoring helpful; 73.56% also indicated that they had a better understanding of fractions after going to the tutoring group. In general, while students found both Cuisenaire Rods and number lines to be helpful, slightly more students found number lines to be more helpful (93.10%) than Cuisenaire Rods (86.21%). Most students (90.80%) indicated that they understood fraction addition well; in contrast, fewer students (54.02%) stated that they understood division of fractions. Fifty-seven percent of the students in the tutoring group indicated that they found fractions difficult; this is somewhat lower than what we had anticipated given their relatively low scores on the screening battery. See Table 17 for a summary of responses from the student survey.

Table 17

Student Appraisal Survey

Survey Item	Percentage of Students		
	Yes, most of the time	*Yes, sometimes*	*No, I did not like going to fractions tutoring*
Did you like going to your fractions tutoring group?	54.02	42.53	2.30
	Yes, helpful	*Yes, sort of helpful*	*No, not helpful*
Did you find fractions tutoring helpful?	85.06	13.79	0.00

Survey Item	Percentage of Students		
	Yes, better	*Yes, sort of better*	*No, not better*
Do you understand fractions better now after going to your tutoring group?	73.56	25.29	0.00
	Yes, difficult	*Yes, sort of difficult*	*No*
Do you find fractions difficult?	3.45	54.02	41.38
	Yes, helpful	*Yes, sort of helpful*	*No, not helpful*
Cuisenaire Rods helped me understand fractions.	47.13	39.08	12.64
Number lines helped me understand fractions.	55.17	37.93	5.75
How well do you know understand these fraction topics?	*I know these well*		
Equivalent fractions	60.92		
Adding fractions	90.80		
Subtracting fractions	79.31		
Multiplying fractions	74.71		
Dividing fractions	54.02		
Putting fractions on a number line	56.32		

Note. Total sample size *N* = 87 intervention students. Some percentages do not sum to 100 because some students did not respond to all survey items.

Tutor survey. All tutors indicated that their students had improved as result of tutoring. See Table 18. All also felt that their students appeared to be more confident and that they were able to articulate their understanding of fraction concepts. Seventy percent of the tutors also indicated that students were writing coherent explanations, understanding and using math vocabulary, and making fewer calculation errors. Only

40% indicated that they noticed students use number lines while solving problems. While most of the tutors (70%) felt that the *TransMath* curriculum was somewhat easy to follow, they acknowledged that the lessons in *TransMath* were very well organized (90%). Most tutors felt that working with five students per group was manageable (30% agree; 60% strongly agree); however, many also felt that it was difficult to meet students' needs as they were at different levels (30% agree; 30% strongly agree). While 80% of the tutors felt that the training provided by the research staff was sufficient, they identified areas where additional training would be useful: (a) more practice with Cuisenaire Rods, (b) strategies for dealing with disruptive students or helping English language learners with written explanations, and (c) examples of what a lesson should look like.

When asked what the most difficult aspect of tutoring was, tutors raised issues such as having limited time to cover the material, having to deal with student behavior issues, catering to students at different levels, having uncooperative classroom teachers, and not having a reliable schedule and/or tutoring room. Tutors also suggested holding tutoring sessions when students are studying fractions in their regular classroom rather than before or after fractions have been taught, conducting tutoring sessions in the morning rather than in the afternoon, having smaller groups (3–4 students) to make sure all students are on track and to provide more individualized instruction, and excluding students with behavioral/emotional issues as they were unable to manage them.

Table 18

Tutor Appraisal Survey

Survey Item	Percentage of Teachers			
	Yes			
Do you think the students in your tutoring group(s) improved?	100.00			
What improvements did you notice in your students?				
Appear more confident	100.00			
Able to articulate their understanding of fraction concepts	100.00			
Coherent written explanations	70.00			
Understand and use math vocabulary	70.00			
Fewer errors in calculations	70.00			
Participate in classroom discussions	50.00			
Use number line while solving problems	40.00			
	Always	*Mostly*	*Somewhat*	*Not at all*
Did you feel the curriculum was easy to follow?	0.00	20.00	70.00	10.00
	Always	*Mostly*	*Somewhat*	*Not at all*
Did you feel the lessons were well organized?	0.00	90.00	10.00	0.00
Was the training provided by the research staff sufficient?	80.00	20.00	0.00	0.00
	Strongly Agree	*Agree*	*Disagree*	*Strongly Disagree*
Working with five students per group was manageable	60.00	30.00	10.00	0.00
It was difficult to meet student needs as they were at different levels	30.00	30.00	30.00	10.00
Student behavior problems took time away from teaching	30.00	40.00	30.00	0.00
It was difficult to complete the lessons in the allocated time	30.00	20.00	40.00	10.00
School staff were helpful and welcoming	40.00	60.00	0.00	0.00
Teachers were friendly and cooperative	30.00	70.00	0.00	0.00
Room provided for tutoring was satisfactory	60.00	30.00	0.00	10.00
	Definitely Yes	*Probably*	*Maybe*	*Definitely No*
Would you tutor again?	50.00	30.00	10.00	10.00
Would you test again?	70.00	10.00	20.00	0.00

Note. Total number of tutors = 10.

***Classroom teacher survey*.** Sixty percent of the classroom teachers indicated that their students (who received the fractions intervention) showed improved fraction understanding. See Table 19. Another 36.67% indicated that their students had improved, but only minimally. Close to 70% of teachers felt that the fractions tutoring was beneficial or very beneficial for their students. When asked to compare the fractions performance of their intervention students with their class peers, 66.67% of the teachers indicated that the fractions performance was average; another 3.33% indicated that it was above average, and another 23.33% indicated that it was below class average. Like the tutors (but to a lesser extent) the teachers indicated that the students appeared more confident (75%), were participating in classroom discussions (59.38%), and were able to articulate their understanding of fractions concepts (46.88%).

Table 19

Classroom Teacher Appraisal Survey

Survey Item	Percentage of Teachers			
	Very beneficial	*Beneficial*	*Somewhat beneficial*	*Not at all beneficial*
Overall, how beneficial was the fractions tutoring for your small group of students?	25.00	43.75	31.25	0.00
	Above class average	*At class average*	*Somewhat below class average*	*Way below class average*
How is their fractions performance when compared to their peers?	3.33	66.67	23.33	6.67
		Yes	*Somewhat improved*	*No*
Have these students shown an improvement in their understanding of fractions?		60.00	36.67	3.33
	Yes			
On average, what improvements relating to fractions have you seen for the students who participated in tutoring:				
Appear more confident	75.00			
Participate in classroom discussions	59.38			
Able to articulate their understanding of fraction concepts	46.88			
Fewer errors in calculations	28.13			
Use number line while solving problems	9.38			
If fractions tutoring were to be offered again next school year, would you allow your students to participate?	78.13			

Note. Total sample size *N* = 30–32 teachers because some teachers did not respond to all survey items.

5. Summary

The purpose of this randomized controlled trial was to assess the impact of a fractions intervention on at-risk fifth-graders' understanding of foundational fractions concepts as well as grade-level competence with fractions operations. The fractions intervention used 52 modified lessons from the *TransMath* curriculum. Each modified lesson was 35 minutes in duration. The intervention was provided 3–4 times a week.

The impact of the fractions intervention was assessed using a randomized controlled trial. Findings indicate that students who received the *TransMath* fractions intervention performed significantly better on a range of fractions outcomes than students who did not receive the intervention. Effect sizes (Hedges' *g*) ranged from .66 to 1.08, and *p*-values were all less than .0001 even after the Benjamini-Hochberg correction.

In addition, a series of five performance assessments were used to assess students' understandings of the underlying concepts and their rationale for solving the problems. On a randomly selected sub-sample, intervention students performed significantly better than comparison students on all five performance assessments (g = .68 to 1.23); the impacts were all statistically significant at p = .01, even after correction for multiple comparison.

These findings demonstrate that a fractions intervention that is grounded in a curriculum that consistently emphasizes use of the number line to build understanding of foundational concepts as well as operations and that emphasizes problem solving and provides opportunities to check for student understanding has the capacity to improve at-risk students' grade-level fractions performance. The findings have a meaningful and practical value, given the well-established, predictive relationship between knowledge of fractions at age 10 (i.e., 5^{th} grade) and performance in algebra and overall mathematics achievement in 11^{th} grade (Siegler et al. (2012). With a solid foundation in fractions and other rational number topics being a key predictor of success in algebra (e.g., Booth, Newton, & Twiss-Garrity, 2014; Geary, Hoard, Nugent, & Bailey, 2012; Siegler et al., 2012), it is critical that students attain proficiency in fractions in upper-elementary grades before moving to more advanced mathematics in middle school.

References

Baldwin, S. A., Bauer, D. J., Stice, E., & Rohde, P. (2011). Evaluating models for partially clustered designs. *Psychological Methods, 16*(2), 149–165. doi:10.1037/a0023464

Baldwin, S. A., Murray, D. M., & Shadish, W. (2005). Empirically supported treatments or type I errors? Problems with the analysis of data from group-administered treatments. *Journal of Consulting and Clinical Psychology, 73*(5), 924–935. doi:10.1037/0022-006X.73.5.924

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. doi:10.1037//0022-3514.51.6.1173

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research, 43*(2), 210–236. doi:10.1080/00273170802034810

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289–300. Retrieved from http://www.jstor.org/stable/2346101

Booth, J. L., Newton, K. J., & Twiss-Garrity, L. K. (2014). The impact of fraction magnitude knowledge on algebra performance and learning. *Journal of Experimental Child Psychology*, *118*, 110–118. doi:10.1016/j.jecp.2013.09.001

California Department of Education. (2013). *Mathematics*. Retrieved from https://www.cde.ca.gov/ci/ma/cf/mathfwchapters.asp

Carpenter, T. P., Fennema, E., Franke, M. L., Levi, L., & Empson, S. B. (2014). *Children’s mathematics: Cognitively guided instruction* (2nd ed.). Portsmouth, NH: Heinemann.

Chval, K., Lannin, J., Jones, D., & Dougherty, B. (2013). *Putting essential understanding of fractions into practice in grades 3–5*. Reston, VA: The National Council of Teachers of Mathematics.

Clarke, B., Doabler, C. T., Smolkowski, K., Kurtz Nelson, E., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a Tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness*, *9*(4), 607–634. doi:10.1080/19345747.2015.1116034

Clarke, D., Roche, A., Sullivan, P., & Cheeseman, J. (2014). Creating a classroom culture that encourages students to persist on cognitively demanding tasks. In K. Karp (Ed). *Annual perspectives in mathematics education: Using research to improve instruction* (pp. 67–76). Reston, VA: NCTM.

Dasgupta, A., Lawson, K. A., & Wilson, J. P. (2010). Evaluating equivalence and noninferiority trials. *American Journal of Health-System Pharmacy, 67*(16), 1337–1343. doi:10.2146/ajhp090507

Fennell, F. (2011). All means all. In F. Fennell (Ed.), *Achieving fluency: Special education and mathematics* (pp. 1–14). Reston, VA: NCTM.

Foorman, B.R., Schatschneider, C., Eakin, M.N., Fletcher, J.M., Moats, L.C., & Francis, D.J. (2006). The impact of instructional practices in grades 1 and 2 on reading and spelling achievement in high poverty schools. *Contemporary Educational Psychology*, *31*(1), 1–29. doi:10.1016/j.cedpsych.2004.11.003

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

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Malone, A. S., Wang, A., . . . 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. doi:10.1086/686303

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology*, *104*(1), 206–223. doi:10.1037/a0025398

Gersten, R., Carnine, D., & Williams, P. (1982). Measuring implementation of a structured educational model in an urban setting: An observational approach. *Educational Evaluation and Policy Analysis, 4*(1), 67–79. doi:10.3102/01623737004001067

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, *79*(3), 1202–1242.

doi:10.3102/0034654309334431

Gersten, R., Rolfhus, E., Clarke, B., Decker, L. E., Wilkins, C., & Dimino, J. (2015). Intervention for first graders with limited number knowledge: Large-scale replication of a randomized controlled trial. *American Educational Research Journal*, *52*(3), 516–546. doi:10.3102/0002831214565787

Gersten, R., Taylor, M. J., Keys, T. D., Rolfhus, E., & Newman-Gonchar, R. (2014). *Summary of research on the effectiveness of math professional development approaches*. (REL 2014–010). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs

Hannan, P. J., & Murray, D. M. (1996). Gauss or Bernoulli?: A Monte Carlo comparison of the performance of the linear mixed-model and the logistic mixed-model analyses in simulated community trials with a dichotomous outcome variable at the individual level. *Evaluation Review, 20*(3), 338–352. doi:10.1177/0193841x9602000306

Houghton Mifflin Harcourt. (2018). *GO math!.* Retrieved from https://www.hmhco.com

Illustrative Mathematics. (2013). *Illustrative mathematics*. Retrieved from http://www.illustrativemathematics.org/

Instructional Research Group [IRG]. (2014). Test for understanding of fractions: 4th grade (TUF-4) [Assessment]. Los Alamitos, CA: Author.

Instructional Research Group [IRG]. (2015). Test for understanding of fractions: 5th grade (TUF-5) [Assessment]. Los Alamitos, CA: Author.

Jayanthi, M., Gersten, R., Taylor, M. J., Smolkowski, K., & Dimino, J. (2017). *Impact of the Developing Mathematical Ideas professional development program on grade 4 students' and teachers' understanding of fractions* (REL 2017–256). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from http://ies.ed.gov/ncee/edlabs

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology*, *116*(1), 45–58. doi:10.1016/j.jecp.2013.02.001

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*(5), 602–619. doi:10.1177/0193841X8100500502

Keppel, G., & Zedeck, S. (1989). *A series of books in psychology. Data analysis for research designs: Analysis of variance and multiple regression/correlation approaches.* New York, NY: Henry Holt & Co.

Kromrey, J. D., & Dickinson, W. B. (1996). Detecting unit of analysis problems in nested designs: Statistical power and type I error rates of the F test for groups-within-treatments effects. *Educational and Psychological Measurement, 56*(2), 215–231. doi:10.1177/0013164496056002003

McGraw-Hill Education. (2017a). *California math*. Retrieved from https://www.mheducation.com

McGraw-Hill Education. (2017b). *My math*. Retrieved from https://www.mheducation.com

Mitchell, M. A., & Maxwell, S. E. (2013). A comparison of the cross-sectional and sequential designs when assessing longitudinal mediation. *Multivariate Behavioral Research*, *48*(3), 301–339. doi:10.1080/00273171.2013.784696

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Muthén, L. K. & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Los Angeles: Muthén & Muthén. Retrieved from https://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf

National Center for Education Statistics (NCES). (2009). *The nation's report card: Mathematics 2009* (NCES 2010–451). Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2009/2010451.pdf

National Governors Association Center for Best Practices (NGA Center), & Council of Chief State School Officers (CCSSO). (2010). *Common core state standards for mathematics*. Washington DC: Author. Retrieved from http://www.corestandards.org/Math

Pearson Education. (2015). *Mathematics practice tests*. Retrieved from https://parcc.pearson.com/practice-tests/math/

Pianta, R. C., Hamre, B. K., & Mintz, S. (2012a). *Classroom assessment scoring system: Upper elementary manual*. Charlottesville, VA: Teachstone.

Pianta, R. C., Hamre, B. K., & Mintz, S. (2012b). *Upper elementary and secondary CLASS technical manual*. Retrieved from http://cdn2.hubspot.net/hubfs/336169/Technical_Manual.pdf

Powell, S. R., & Fuchs, L. S. (2015). Intensive intervention in mathematics. *Learning Disabilities Research & Practice*, *30*(4), 182–192. doi:10.1111/ldrp.12087

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891. doi:10.3758/BRM.40.3.879

Roberts, C., & Roberts, S. A. (2005). Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials, 2*(2), 152–162. doi:10.1191/1740774505cn076oa

SAS Institute. (2016). SAS/STAT® (Version 14.2) [user's guide]. Cary, NC: SAS Institute. Retrieved from http://support.sas.com/documentation/index.html

Schumacher, R., Jayanthi, M., Gersten, R., Dimino, J., Spallone, S., & Haymond, K. (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. doi:10.1111/ldrp.12169

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworkth, K., Claessens, A., Engel, M., . . . Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691–697. doi:10.1177/0956797612440101

Siegler, R. S., & Pyke, A. A. (2013). Developmental and individual differences in understanding of fractions. *Developmental Psychology*, *49*(10), 1994–2004. doi:10.1037/a0031200

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*(4), 273–296. doi:10.1016/j.cogpsych.2011.03.001

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. doi:10.1177/0956797611417632

Smith, T. M., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating math recovery assessing the causal impact of a diagnostic tutoring program on student achievement. *American Educational Research Journal, 50*(2), 397–428. doi:10.3102/0002831212469045

Spector, P. E., & Brannick, M. T. (2011). Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods*, *14*(2), 287–305. doi:10.1177/1094428110369842

U.S. Department of Education [U.S. ED], Institute of Education Sciences [IES], & What Works Clearinghouse [WWC]. (2017). *What Works Clearinghouse: Procedures and standards handbook* (Version 4.0). Retrieved from http://whatworks.ed.gov van Belle, G. (2008). *Statistical rules of thumb* (2nd ed.). New Jersey, NJ: Jon Wiley & sons.

von Soest, T., & Hagtvet, K. A. (2011). Mediation analysis in a latent growth curve modeling framework. *Structural Equation Modeling*, *18*(2), 289–314. doi:10.1080/10705511.2011.557344

Wilkinson, G. S., & Robertson, G. J. (2006). *WRAT4: Wide range achievement test*. Bloomington, MN: NCS Pearson, Inc.

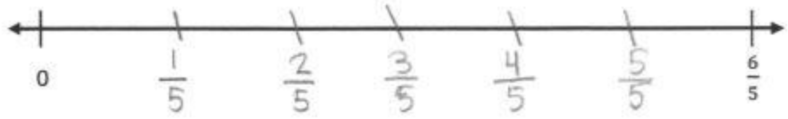
Willett, J. B. (1988). Questions and answers in the measurement of change. *Review of Research in Education*, *15*(1), 345–422. doi:10.3102/0091732X015001345

Woodward, J., & Stroh, M. (2015). *TransMath: Fluency with rational numbers* (3rd ed.). Dallas, TX: Voyager Sopris Learning.

Appendix A

Figure A1

Performance Assessment 1 Intervention Student Response



How did you know where to place $\frac{2}{5}$ on the number line? Explain your thinking.

I knew were $\frac{2}{5}$ went on the number line because you have to have fair shair or elese if will not be equal. $\frac{2}{5}$ go's right after $\frac{1}{5}$ because after 1 is 2 so you can pridict that $\frac{2}{5}$ goes right after $\frac{1}{5}$ with fair share.

Figure A2

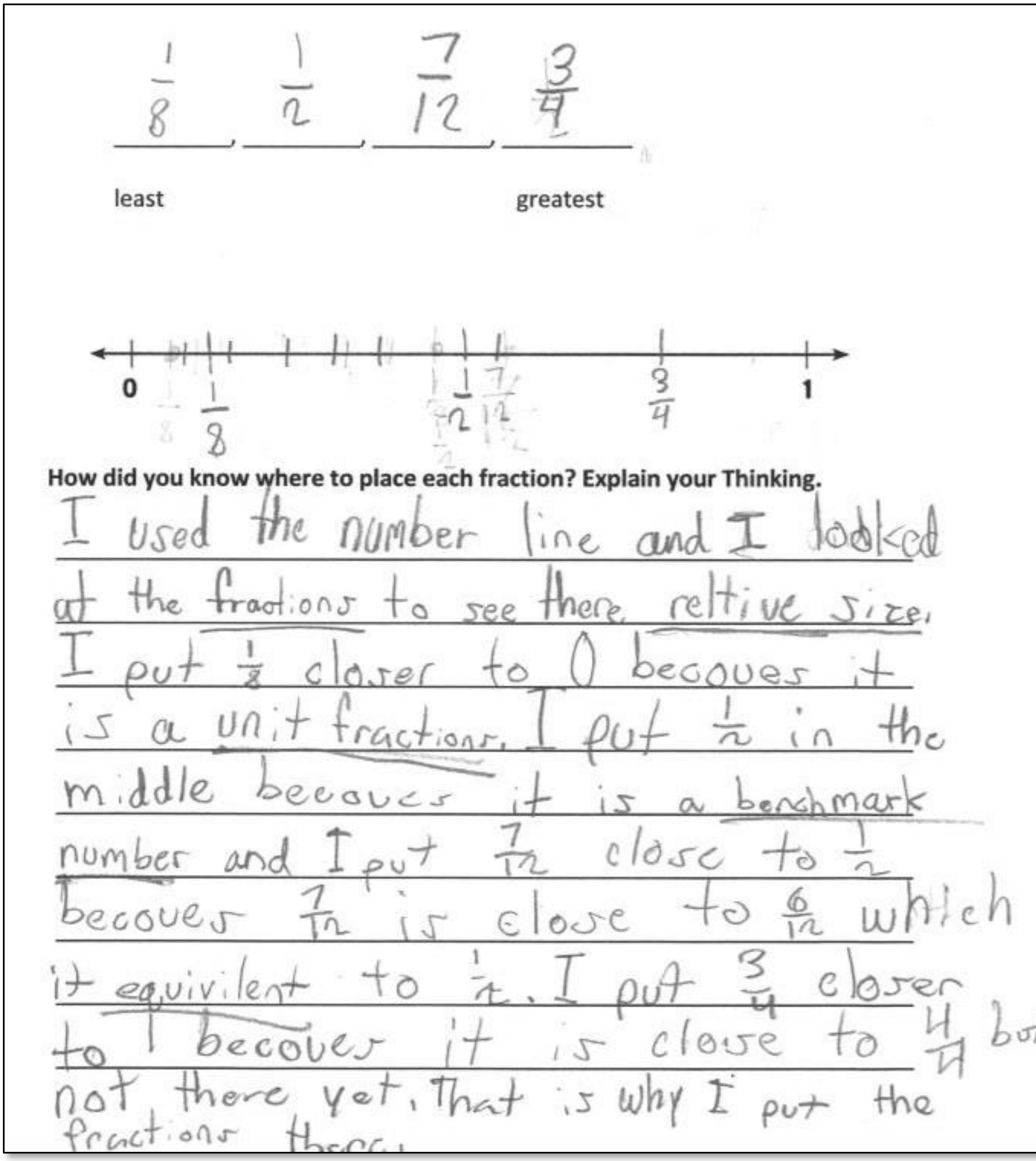
Performance Assessment 1 Comparison Student Response

How did you know where to place $\frac{2}{5}$ on the number line? Explain your thinking.

becuse i did 2 loops then i Knew that $\frac{2}{5}$ was there.

Figure A3

Performance Assessment 2 Intervention Student Response



$\frac{1}{8}$, $\frac{1}{2}$, $\frac{7}{12}$, $\frac{3}{4}$

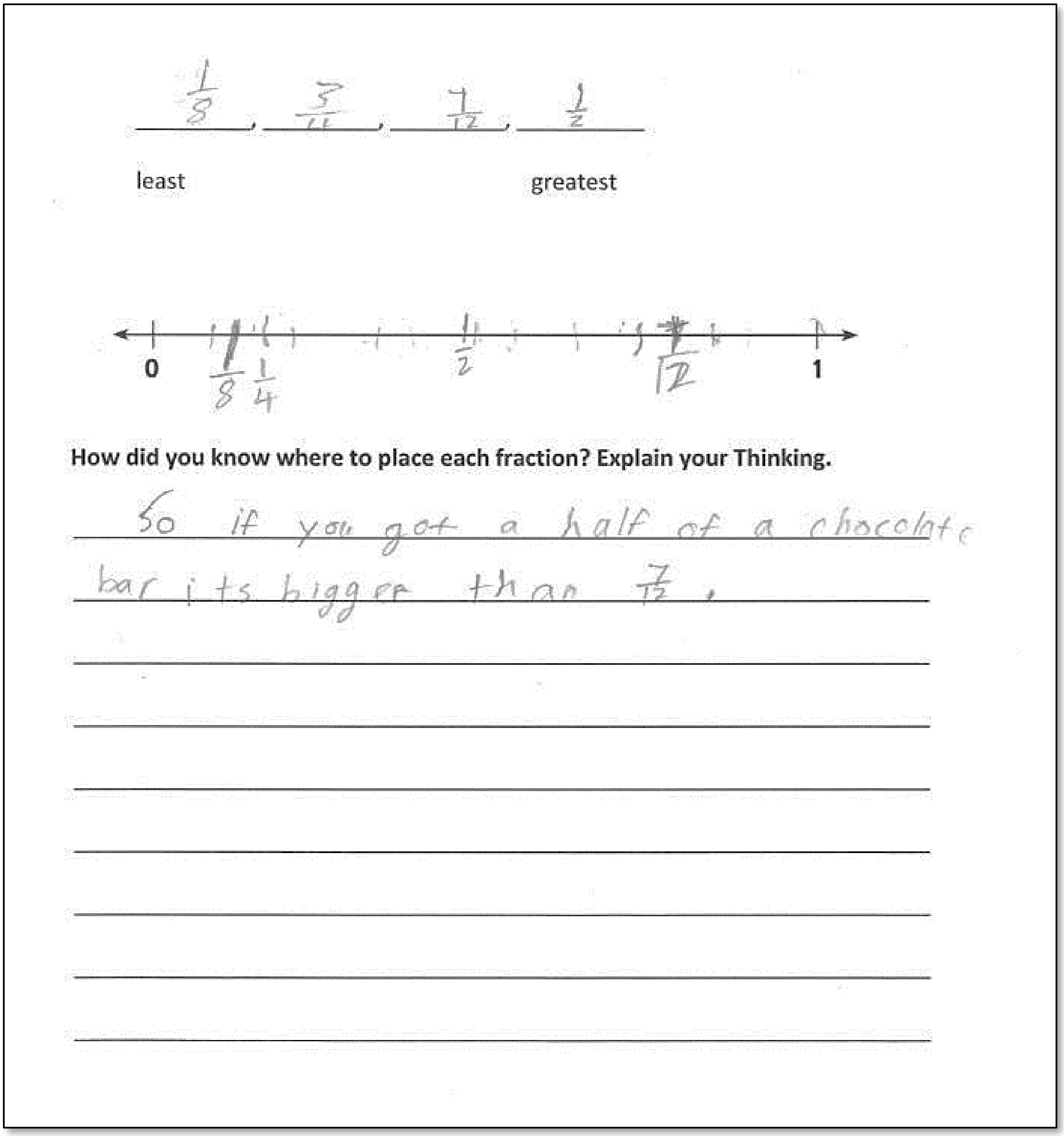
least greatest

How did you know where to place each fraction? Explain your Thinking.

I used the number line and I lookked at the fractions to see there reltive size. I put $\frac{1}{8}$ closer to 0 becoues it is a unit fractions. I put $\frac{1}{2}$ in the middle becoues it is a benchmark number and I put $\frac{7}{12}$ close to $\frac{1}{2}$ becoues $\frac{7}{12}$ is close to $\frac{6}{12}$ which it equivilent to $\frac{1}{2}$. I put $\frac{3}{4}$ closer to 1 becoues it is close to $\frac{4}{4}$ but not there yet. That is why I put the fractions there.

Figure A4

Performance Assessment 2 Comparison Student Response



$\frac{1}{8}$, $\frac{3}{4}$, $\frac{7}{12}$, $\frac{1}{2}$

least greatest

0 $\frac{1}{8}$ $\frac{1}{4}$ $\frac{1}{2}$ $\frac{7}{12}$ 1

How did you know where to place each fraction? Explain your Thinking.

So if you got a half of a chocolate bar its bigger than $\frac{7}{12}$.

Figure A5

Performance Assessment 3 Intervention Student Response

$$\frac{1}{2}+\frac{4}{10}+\frac{7}{10} \qquad \left(\frac{5}{10}+\frac{4}{10}\right)+\frac{7}{10}$$

$$\frac{5}{10}+\frac{4}{10}+\frac{7}{10}= \qquad \frac{9}{10}+\frac{7}{10}=\frac{16}{10} \text{ or } 1\frac{6}{12}$$

Explain your thinking.

Jessie walked $1\frac{6}{12}$ of a mile on Saturday. I found that out by changing the denominator on $\frac{1}{2}$ because two fractions had the same denominator and $\frac{1}{2}$ didn't. So I changed $\frac{1}{2}+\frac{5}{10}$ and added $\frac{5}{10}+\frac{4}{10}+\frac{7}{10}$. First, I put $\frac{5}{10}$ and $\frac{4}{10}$ in parenthesses and got $\frac{9}{10}$. Then, I added $\frac{9}{10}$ and $\frac{7}{10}$ together and got $\frac{16}{10}$ or $1\frac{6}{12}$. My final answer was $\frac{16}{10}$ or $1\frac{6}{12}$ which was improper so I simplifyed it and got $1\frac{6}{12}$ which is a mixed number. Thats how I knew Jessie walked $1\frac{6}{12}$ of a mile.

Figure A6

Performance Assessment 3 Comparison Student Response

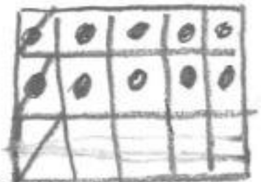
Explain your thinking.

AS I was makeing the modle I Put $\frac{7}{10}$ and $\frac{1}{2}$ hafe together and got $\frac{6}{10}$ so I Put that as MY ahser.

Figure A7

Performance Assessment 4 Intervention Student Response

$$\frac{1}{5} \cdot \frac{2}{3} = \frac{2}{15}$$

 $= \frac{2}{15}$

Explain your thinking.

I multiplied $\frac{1}{5}$ and $\frac{2}{3}$ becoves in the question it says what fraction of her set of Legos did she use to build the fire truck. What I did was multiple across and I used a area model to check. I multiplied 1 and 2 and got 2 for my numerator and multiplied 5 and 3 and got 15 for the denominator. That was how I got $\frac{2}{15}$ from multiping and from the area modol.

Figure A8

Performance Assessment 4 Comparison Student Response

$\frac{2}{3} - \frac{1}{5}$

3: 3, 6, 9, 12, 15

5: 5, 10, 15

$\frac{2^{10}}{15} - \frac{1^{3}}{15}$

$\frac{7}{15}$

Explain your thinking.

I got my awencer by subtracting fraction.

Figure A9

Performance Assessment 5 Intervention Student Response

$$\overset{9}{\cancel{10}}\ \frac{14}{10} + \frac{10}{10}$$

$$-\ 7\,\frac{9}{10}$$

$$2\,\frac{5}{10} \div \frac{2}{2} = 2\,\frac{1}{2}$$

$$\frac{2}{5} \cdot \frac{2}{2} = \frac{4}{10}$$

Explain your thinking.

First I looked if the numbers are lined up. Then I checked if I needed to aldter, which I need to. So $\frac{2}{5}$ has a common denominator with 2, so I got the answer $\frac{4}{10}$. But I still can't subtract because 2 can not subtract with the fraction $\frac{9}{10}$ so you have to cross out the whole number then burrow one then turn it into a whole number $\frac{10}{10}$. Finally add 14 and 10. Lastly preform which is to subtract. Then I got the answer $2\frac{5}{10}$. Finally check if need to simplify. Wich we need to then got the finally answer $2\frac{1}{2}$.

Figure A10

Performance Assessment 5 Comparison Student Response

$10\frac{2}{5} \times 7\frac{9}{10} = 70\frac{18}{10}$

$1\frac{8}{10}$

R8

$10\overline{)18}$ = 1

-10

08

Explain your thinking.

I learned that when you have a inproper fraction you should divided the number. My answer is $1\frac{8}{10}$.

Appendix B

Figure B1

Curriculum-aligned Posttest Moderated by WRAT-4 Pretest

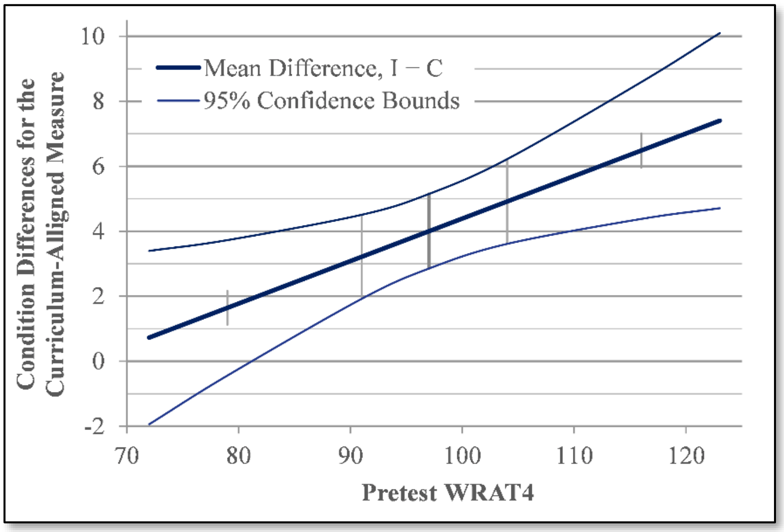


Figure B2

TUF-5 Posttest Moderated by WRAT-4 Pretest

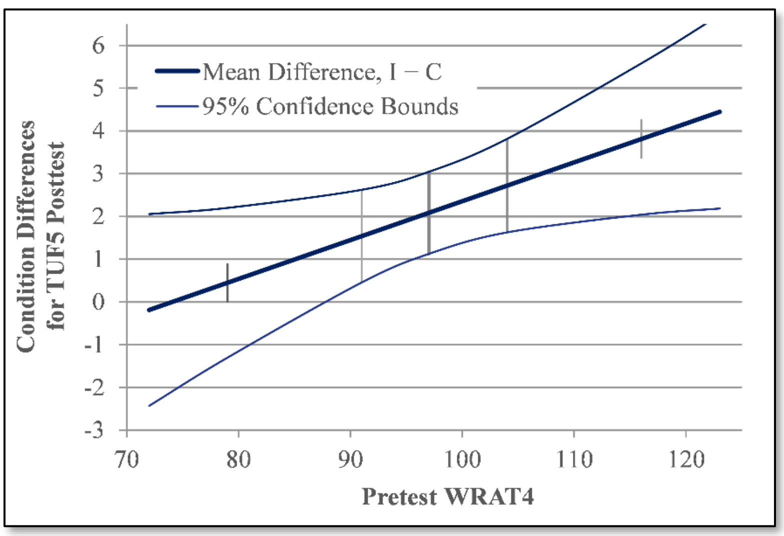
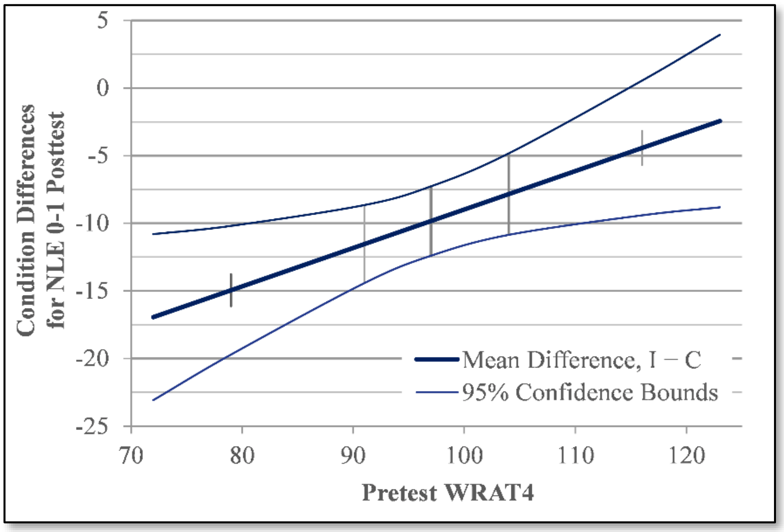


Figure B3

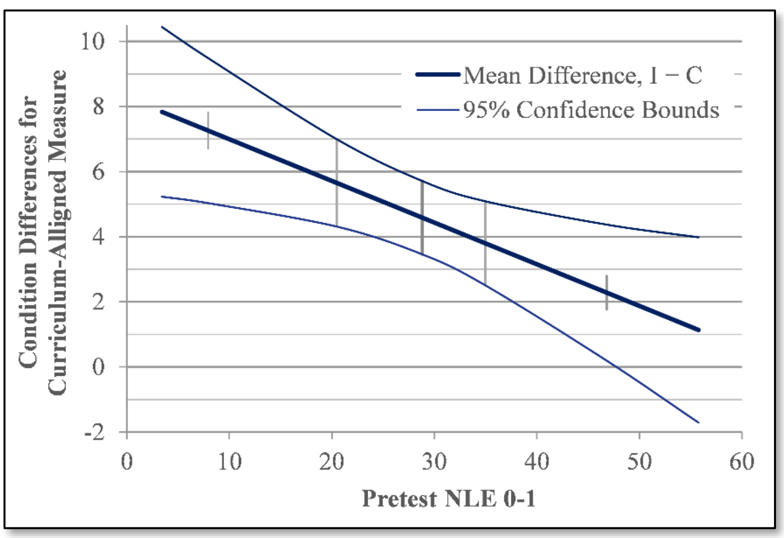
NLE 0–1 Posttest Moderated by WRAT-4 Pretest



Note. *NLE 0–1* and *NLE 0–2* are reported in terms of percent absolute *error* (i.e., a *low* score indicates *high* performance).

Figure B4

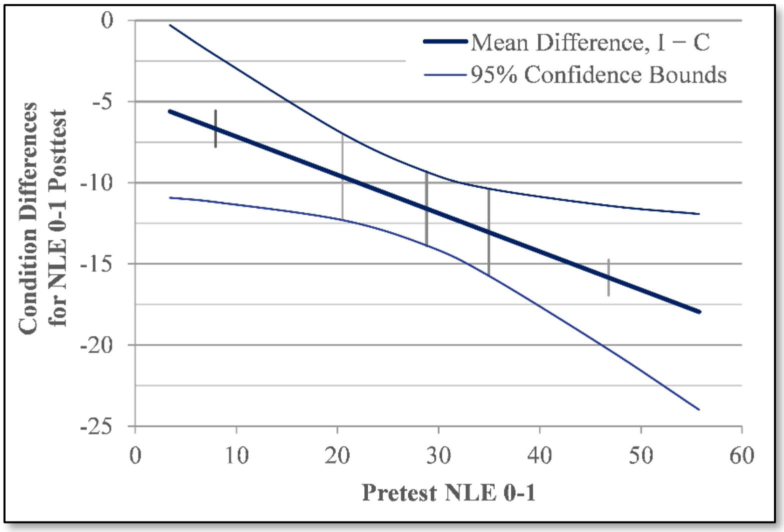
Curriculum-aligned Posttest Moderated by NLE 0–1 Pretest



Note. *NLE 0–1* and *NLE 0–2* are reported in terms of percent absolute *error* (i.e., a *low* score indicates *high* performance).

Figure B5

NLE 0–1 moderated by NLE 0–1 pretest



Note. *NLE 0–1* and *NLE 0–2* are reported in terms of percent absolute *error* (i.e., a *low* score indicates *high* performance).

Table B1

NLE 0–1 Posttest Moderated by District 2 vs. Not (Districts 1 and 3)

Effect	Estimate	Lower	Upper	*t (df)*	*p*
I-C District2	−13.902	−17.662	−10.142	−7.34	< .0001
I-C Not	−7.476	−11.216	−3.736	−3.96	.001
I-C District2–Not	−6.426	−11.730	−1.123	−2.40	.0181

Note. Total Analytic Sample *N* = 186 students (I = 87, C = 99). *NLE 0–1* and *NLE 0–2* are reported in terms of percent absolute *error* (i.e., a *low* score indicates *high* performance).

Table B2

NLE 0–1 Posttest Moderated by District 3 vs. Not (Districts 1 and 2)

Effect	Estimate	Lower	Upper	*t (df)*	*p*
I-C District3	−4.289	−10.190	1.612	−1.44	.1526
I-C Not	−12.001	−14.828	−9.174	−8.42	< .0001
I-C District3–Not	7.713	1.170	14.256	2.34	.0213

Note. Total Analytic Sample *N* = 186 students (I = 87, C = 99). *NLE 0–1* and *NLE 0–2* are reported in terms of percent absolute *error* (i.e., a *low* score indicates *high* performance).

Table B3

TUF-4 Posttest Moderated by Free and Reduced-Price Lunch (FRL) vs. Not

Effect	Estimate	Lower	Upper	*t (df)*	*p*
I-C FRL	2.103	0.583	3.622	2.77	.0075
I-C Not	4.884	2.763	7.006	4.59	< .0001
I-C FRL–Not	−2.782	−5.399	−0.165	−2.11	.0374

Note. Total Analytic Sample *N* = 186 students (I = 87, C = 99).

Table B4

NLE 0–1 Posttest Moderated by Free and Reduced-Price Lunch (FRL) vs. Not

Effect	Estimate	Lower	Upper	*t (df)*	*p*
I-C FRL	−12.418	−15.866	−8.970	−7.14	< .0001
I-C Not	−6.231	−10.965	−1.496	−2.61	.0104
I-C FRL–Not	−6.188	−12.087	−0.288	−2.08	.0400

Note. Total Analytic Sample *N* = 186 students (I = 87, C = 99). *NLE 0–1* and *NLE 0–2* are reported in terms of percent absolute *error* (i.e., a *low* score indicates *high* performance).