Regional Educational Laboratory Midwest Teaching Fractions Toolkit Evaluation

OMB# 1850-NEW

Supporting Justification
for OMB Clearance
Section B

Submitted by:

National Center for Education Evaluation (NCEE)
Institute of Education Sciences (IES)
U.S. Department of Education
Washington, DC

May 2023

Contents

Overview

The U.S. Department of Education (ED), through its Institute of Education Sciences (IES), requests clearance for the recruitment materials and data collection protocols under the Office of Management and Budget (OMB) clearance agreement (OMB Number 1850-NEW) for activities related to the Regional Educational Laboratory (REL) REL Midwest Program under contract 91990022C0011.

Computational skills with fractions underpin advanced mathematics (Booth & Newton, 2012), are essential for success in high school mathematics, and are a prerequisite for college-level mathematics courses (Siegler & Lortie-Forgues, 2015). Unfortunately, student difficulty with fractions is well documented (Barbieri et al., 2020; Liu, 2018; Siegler & Lortie-Forgues, 2015). Even after studying fractions and related topics for several years, U.S. students often lack a conceptual understanding of fractions (Siegler et al., 2010). These fraction difficulties are widespread and critical to address because "early fraction knowledge strongly predicts later mathematics knowledge even after children's IQ, reading comprehension, working memory, whole-number arithmetic knowledge, race, ethnicity, and parental education and income are statistically controlled" (Fazio et al., 2016, p. 1).

Difficulties with fractions-related content are not confined to students; teachers often have difficulties as well. Teachers often struggle with fraction computation (Harvey, 2012), and many practicing and preservice teachers have considerable difficulty with fraction operations, including multiplication and division (Tekin-Sitrava, 2020; Whitehead & Walkowiak, 2017). In a recent study, only 42 percent of prospective teachers who attempted to solve equations with fractions solved the equations correctly (Jones et al., 2020). Although teachers' work with students is at the heart of student learning, administrators also are essential in building systemic approaches to improving teaching and learning and in providing the appropriate supports for teacher success (Park et al., 2019). Therefore, administrators need to be prepared to set standards, identify needs, and provide the appropriate supports if teachers are to be effective.

To address these needs, REL Midwest is developing a toolkit (the Teaching Fractions Toolkit) that supports teachers to enact evidence-based practices summarized in *Developing Effective Fractions Instruction for Kindergarten Through 8th Grade* (Siegler et al., 2010). Drawing on the recommendations and implementation steps outlined in the practice guide, the toolkit will address teacher understanding of fraction computation, rates, and ratios, as well as implications for classroom practiced related to fractions content for grade 6 teachers. REL Midwest is developing the toolkit in collaboration with district partners in Illinois.

ED, in consultation with the American Institutes for Research® (AIR®), is planning a two-part evaluation of the toolkit in 40 Illinois public schools across 6–10 school districts. The evaluation will consist of an impact study and an implementation study. OMB approval is being requested for a multimode data collection and analysis of a group of schools, students, and staff members in these Illinois public schools.

Description of the Teaching Fractions Toolkit

The Teaching Fractions Toolkit is based on and supports implementation of five evidence-based recommendations in the What Works Clearinghouse practice guide *Developing Effective Fractions Instruction for Kindergarten Through 8th Grade* (Siegler et al., 2010). The practice guide recommendations (see Box 1) are based on rigorous research for improving K–8 students' understanding of fractions, with the expectation that general education teachers, mathematics specialists and coaches, special educators, and administrators will use these resources to improve their teaching of fractions.

This toolkit includes two types of supports: teacher supports and institutionalizing supports for administrators and mathematics leaders who support mathematics teachers.

The primary audience for the teacher supports is grade 6 mathematics teachers in general education classrooms. Teacher supports include six teacher professional development (PD) modules. Each professional development module consists of two synchronous sessions led by a PD facilitator, separated by approximately three hours of asynchronous assignments in the interim between sessions. In each module, teachers engage in individual and collaborative PD activities, including exploration of mathematics tasks, student work analysis, lesson planning, the use of formative assessment items, and reflection on classroom practice, all of which will support teachers' understanding related to the implementation steps for practice guide Recommendations 2–4 as well as how to mitigate possible roadblocks identified for Recommendations 2–4. The toolkit also includes associated resources to support engagement in PD in each module, including mathematics tasks, interactive applets, protocols for student work analysis and planning, videos, student artifacts, readings, and reflection prompts. The toolkit includes a teacher reflection tool to assess initial and developing classroom practices aligned with the practice guide recommendations and questions to inform lesson planning and reflection. The teacher PD is designed so that it can be used with in-person meetings or fully online for all activities (synchronous and asynchronous). The modules and materials will be designed with flexibility so that local facilitators and teachers will be able to implement all or part of the PD in an in-person environment if they choose to do that. The guidance for facilitators will make suggestions about how to lead teacher discussions either in person, if feasible, or via videoconference using whatever videoconference platform the district employs.

> **Box 1: Recommendations in the *Developing Effective Fractions Instruction for Kindergarten Through 8th Grade* practice guide**
>
> 1. Build on students' informal understanding of sharing and proportionality to develop initial fraction concepts.
>
> 2. Help students recognize that fractions are numbers and that they expand the number system beyond whole numbers.
>
> 3. Help students understand why procedures for computations with fractions make sense.
>
> 4. Develop students' conceptual understanding of strategies for solving ratio, rate, and proportion problems before exposing them to cross-multiplication as a procedure to use to solve such problems.
>
> 5. Professional development programs should place a high priority on improving teachers' understanding of fractions and of how to teach them.

The primary audience for the institutionalizing supports is administrators and mathematics leaders (principals, assistant superintendents, curriculum directors, mathematics coaches, and teacher leaders) who support teachers of mathematics. Institutionalizing supports include

- Three videos—one to introduce the toolkit and two to introduce what the practice guide recommendations look like in practice
- Two leader handouts—one summarizing the practice guide recommendations and one outlining the progression of fraction content represented in the practice guide
- A tool for administrators and leaders to assess district conditions to support fractions instruction
- Facilitation guides for school leaders to lead professional development for grade 6 teachers

The institutionalizing supports will bolster the understanding of administrators and mathematics leaders of the importance of the mathematics content embodied in the practice guide recommendations; inform them about the research basis for teacher practices included in the recommendations; guide decisions about supporting teachers to enact the recommendations; and support leaders such as mathematics coaches or other PD providers to lead the PD that is part of the teacher supports.

All materials that users need in order to implement the teacher PD and other toolkit activities and supports are included in the toolkit and will be accessible in one central online location (https://ies.ed.gov/ncee/rel/Midwest/Toolkit) with a clear and user-friendly linked menu on the landing page. The toolkit development team will work with the IES website contractor to get the online platform ready prior to the start of the evaluation so that participating educators will be able to access all toolkit materials online. The landing page will have a brief overview of the toolkit resources, environment, and overarching goals plus sections for institutionalizing supports and teacher supports. The teacher supports section includes the six PD modules. Each module will include a participant workbook, a facilitator guide, and two slides decks (one for each synchronous meeting). Modules will be linked for easy cross-movement and include a navigation menu to the module overview, learning objectives, individual learning activities, a link to resources and tools for that module, the teacher practice monitoring tool, support tips, a glossary of terms and acronyms, and references. All resources will be navigable with a screen reader. When clicked, links will appear in a new tab or window so that the user remains connected to the module. Videos and animations will be captioned with audio available in transcripts to ensure accessibility and Section 508 compliance. Templates, checklists, and tools will be provided in HTML, PDF, and editable document formats. The modules will include links to some interactive GeoGebra applets for use by teachers and their students when working on mathematics tasks. These applets will be developed in the open-source GeoGebra website and made available to teachers through links from the Teaching Fractions Toolkit website and modules.

Research Questions for the Proposed Evaluation

Data collected for this evaluation will be used to examine the implementation of the toolkit in participating schools and the toolkit's efficacy in improving teacher self-efficacy and practices for fraction computation and rate and ratio instruction, as well as student learning

outcomes in grade 6 mathematics. The impact and implementation research questions (RQs) addressed in this study include the following:

1. What is the impact of the toolkit on grade 6 teachers' self-efficacy and teaching practices for fraction computation and rate and ratio instruction compared to the business-as-usual condition?

2. What is the impact of the toolkit on grade 6 students' performance in solving fraction computation and rate and ratio problems compared to the business-as-usual condition?

3. How did the professional development supports and resources available to grade 6 math teachers differ in treatment and control schools?

4. To what extent is the toolkit implemented with fidelity within each participating school and overall across all participating schools?

5. To what extent is the fidelity of implementation associated with teacher self-efficacy and practices and students' performance on solving fraction computation and rate and ratio problems?

6. What contextual factors support or hinder the adoption and implementation of the toolkit?

7. To what extent do the participating teachers and school leaders perceive the toolkit as usable, useful, and feasible to implement? What aspects of the toolkit do they perceive could be improved?

B1. Respondent Universe and Sample Design

The evaluation team aims to recruit 40 schools in Illinois so that the study will be powered to detect effects of the toolkit on student learning and teacher practice outcomes that are of statistical and practical significance and are comparable in magnitude to those effects reported in previous studies of similar interventions. Because the study will not employ random sampling of districts or schools, districts and schools will be recruited and screened based on the characteristics required by the study design.

The team will restrict the universe of schools to public, non-charter schools that serve students in grade 6. However, schools that have participated in the toolkit development stage will not be eligible for the evaluation. The evaluation team will prioritize outreach to under-resourced districts (e.g., districts that serve large percentages of students from families with low incomes, rural districts) because we expect students and teachers in under-resourced schools are in higher need of and are more likely to benefit from supports and resources provided by the toolkit. The team also will aim to recruit districts from diverse settings in terms of geographic locale and district size.

The team expects to recruit 6–10 districts to participate. Districts that are interested in participating in the study will be asked to complete an online form to provide information to the evaluation team to help determine their eligibility for the study. Districts will be eligible to participate if they serve students in grade 6, are willing to participate in a randomized controlled trial (RCT) with delayed implementation for control schools, and are not already providing professional development in grade 6 math instruction that is of the same type and level of intensity as that is being provided by the toolkit.

The evaluation will employ an experimental design in which schools that are eligible and have agreed to participate in the evaluation will be randomly assigned within blocks to treatment condition (toolkit) or business as usual (control) in summer 2024. Each district with multiple schools participating in the study will serve as its own randomization block. Schools from districts in which only one school is participating in the study will be grouped into blocks based on school locale and prior-year school performance. Within each block, the same number of schools will be assigned to each condition (blocks may differ by one school if an odd number of schools are in the block). For blocks with an odd number of schools, having one additional treatment school (e.g., four schools in the treatment group and three schools in the control group) will be equally as likely as one additional control school (e.g., three schools in the treatment group and four schools in the control group). Hence, an individual school's probability of receiving the intervention will always be 50 percent. In schools assigned to the toolkit group, grade 6 teachers and their administrators will be invited to use the toolkit materials with the guidance of a local facilitator. In control schools, grade 6 teachers and administrators will not have access to the toolkit until after the study.

Within schools, the teacher sample will include teachers who teach at least one regular grade 6 math class. For this evaluation, a *regular grade 6 math class* refers to a class that is designated by the school as a general education class and that teaches the district's middle-track grade 6 math curriculum. This definition excludes advanced classes, such as gifted and talented programs and accelerated classes, as well as remedial classes and self-contained special education classes.

The student sample will include students in regular grade 6 math classes in participating schools. Students in special education and in a self-contained setting will not be included because those students likely will be learning content below grade level. Similarly, students in advanced math classes, such as gifted and talented programs, will not be included because they will be learning content that is above grade level. Students who repeat grade 6 will not be included in the evaluation either because their pretest scores from the prior year would be different from students who do not repeat grade 6.

Table B1 shows the target sample sizes and expected response rates for each level of data collection.

Table B1. Target sample size and anticipated response rate for each level of data collection

Level of sample	Target sample size	Response rate
District	6–10	100%
School (school leader)	40	90%
Teacher	134	85%
Student	2,400	85%

B2. Information Collection Procedures

a. Notification of the Sample and Recruitment

The evaluation team will work with partners at the Illinois State Board of Education and leverage existing relationships with Illinois districts to help widely distribute information about the study to districts across the state. Districts that are interested in participating in the study will be asked to complete an online form to indicate their interest and provide information to the evaluation team to help determine their eligibility for the study. The evaluation team will schedule initial virtual informational meetings with districts that have expressed interest to confirm their interest and eligibility and to answer any questions district leaders may have. At this meeting, the evaluation team will inform district leaders about the roles, responsibilities, and benefits of the study. If district leaders are interested in participating, the evaluation team will ask for their help contacting schools and their ideas for how the study might be a fit for their schools. The team will follow up with one-on-one meetings with school leaders, if requested, to answer questions and confirm their interest. If more than 40 schools agree to participate, the team will randomly select schools to participate in the evaluation. Researchers on the team will ask school districts to sign a memorandum of understanding, indicating that they understand the intervention and the study and that schools will participate in the study regardless of the condition to which they are assigned.

Upon district agreement, the team will reach out to school principals and offer to schedule a school-specific information meeting to provide information directly to teachers and facilitators and to hear their thoughts. The evaluation team will prepare a study information sheet for teachers and distribute it to teachers before the meeting. The team will remain flexible and adaptive in the face of emerging recruitment experiences (e.g., by extending the information session to address any immediate concerns of teachers). The evaluation team will collect consent forms from all eligible teachers in participating schools in late summer 2024, after randomization of schools and prior to the start of the 2024/25 school year. Only those teachers who have consented will participate in data collection for the evaluation.

The student sample will include students taught by teachers who have agreed to participate in the study. The evaluation team will collect informed consent from students' parents or caregivers through either an active consent form or a passive consent (opt-out) form, depending on district or school policy. Only students with parent or caregiver consent will be included in the evaluation.

b. Statistical Methodology for Stratification and Sample Selection

Districts and schools that have expressed interest in participating in the study will be vetted for eligibility. If more than 40 schools agree to participate, the team will randomly select schools to participate in the evaluation.

The evaluation team will conduct school-level random assignment within blocks. Research has shown that blocking often improves the precision of impact estimates but needs to be applied thoughtfully (Pashley & Miratrix, 2022). Pashley and Miratrix (2022) advise forming blocks out of covariates predictive of outcome, keeping the proportion of units treated similar across blocks and analyzing the data properly as a blocked experiment. Each district with

multiple schools participating in the study will serve as its own randomization block. Schools from districts in which only one school is participating in the study will be grouped into blocks based on school locale and prior-year school performance. Within each block, the same number of schools will be assigned to each condition (blocks may differ by one school if an odd number of schools are in the block). For blocks with an odd number of schools, having one additional treatment school (e.g., four schools in the treatment group and three schools in the control group) will be equally as likely as one additional control school (e.g., three schools in the treatment group and four schools in the control group). Hence, an individual school's probability of receiving the intervention will always be 50 percent.

Random assignment will be conducted in summer 2024. Plans for random assignment will be communicated with district and school officials early in the recruitment process to ensure buy-in, and randomization assignments will be carefully documented. To maintain the integrity of the random assignment, all analysis of data will account for these procedures, as described below.

c. Estimation Procedures

Impact analysis (RQs 1 and 2). The impact analyses will be intent-to-treat (ITT) analyses that estimate the impact of the toolkit on teachers of regular grade 6 math classes and their students in the study schools. The basic strategy for the impact analysis is to estimate the difference in outcomes between the intervention and comparison groups, adjusting for the blocking used in random assignment and for person- and school-level covariates. The study will use hierarchical linear modeling to estimate the treatment effect on the student- or teacher-level outcomes of interest. In all analyses, students or teachers are the level 1 unit, and schools are the level 2 unit. The student-, teacher-, and school-level variables expected to be correlated with the outcomes will be used as covariates in all analytic models to improve the precision of the impact estimates and to guard against any bias due to imbalance in baseline covariates that arises due to random chance. The evaluation team will specify the following two-level model:

(1a) $\qquad Y_{ij} = b_{0j} + b_{1j} Pretest_{ij} + b_{2j} X_{ij} + \varepsilon_{ij}$

(2a) $\qquad b_{0j} = g_{00} + g_{01} TRT_j + g_{02} W_j + g_{03} Block_j + u_{0j}$

(2b) $\qquad b_{1j} = g_{10}$

(2c) $\qquad b_{2j} = g_{20}$

where Y_{ij} is the outcome score for student or teacher i within school j, $Pretest_{ij}$ is a pretest score on the measure (when available) for student or teacher i within school j, and X_{ij} represents a vector of individual-level covariates; ε_{ij} is a random term. The ITT estimate is g_{01} in the first equation of the level 2 model (equation 2a). TRT_j is an indicator variable that takes a value of 1 for treatment schools and 0 for control schools; W_j is a vector of school-level covariates; and $Block_j$ represents a series of dummy variables indicating the randomization block of each school. b_{1j} and b_{2j} are coefficients for the pretest measure and individual-level covariates, which are assumed to be the same across schools (g_{10} and g_{20} in equations 2b and 2c).

Our impact models will not analyze students and teachers with missing data on the outcome or covariates.

The evaluation team will conduct exploratory analyses to examine whether the impact of the toolkit on student and teacher outcomes is moderated by student, teacher, and school characteristics (e.g., locale). The analyses will be conducted by incorporating appropriate interaction terms into the main impact models. When a significant interaction is identified, the treatment effect within each group will be presented. Potential student-level moderators include multilingual language learner status, eligibility for the National School Lunch Program, and prior achievement. Potential teacher moderators include teacher experience, class size, and class average prior achievement. Potential school-level moderators include size and locale.

Analysis of service contrast (RQ 3). To provide further context for the impact findings, the evaluation team will analyze the contrast between the math professional development received by teachers in the treatment and control schools in 2024/25, based on data from the teacher survey. The analyses of service contrast also will be based on a two-level model controlling for random assignment block.

Analysis of implementation fidelity (RQs 4 and 5). Fidelity of implementation (RQ 4) will be measured for each of the two toolkit components (institutionalizing supports and teacher supports) over the entire intervention sample ($n = 20$ schools). For each component, the evaluation team will work with the toolkit development team to identify quantifiable implementation indicators for the key activities in the logic model and to set the expectations (or thresholds) for determining whether each component has been implemented with fidelity. For each indicator we will specify the unit at which the indicator is measured (teacher or school), the data sources that will be used to measure that indicator, and the approach to scoring. The indicators and thresholds will help operationalize the logic model and ground the evaluation activities in a common understanding of program expectations. For indicators measured at the teacher level, we will roll up the teacher score to create a school-level score. We will then summarize the school-level indicator scores within each component into a total component score for each school.

In addition, the evaluation team will examine the extent to which the level of implementation is related to the size of the impact (RQ 5). The relationship can be estimated with two-level models similar to those for the ITT analysis presented previously, with the treatment indicator in equation 2a (TRT_j) replaced by a predictor indicating the level of implementation at the school level ($Implementation_j$). This analysis will be limited to the sample of treatment schools.

Analysis of participant experience with implementation (RQs 6 and 7). The evaluation team will analyze data from the teacher survey and interviews of teachers and leaders to understand participants' experiences with implementing the toolkit. The evaluation team will use descriptive statistics (frequencies, means, and standard deviations [SDs]) to analyze teachers' responses to relevant items. Interviews will be analyzed using a Miles and Huberman (1994) approach that utilizes inductive and deductive analyses. The team will first employ descriptive coding and assign codes based on the research and interview protocol

questions. The evaluation team will analyze the interviews using NVivo qualitative software, mapping them onto a coding structure that aligns with the topics covered in the interview protocols. The team will analyze a subset (20 percent) of interviews to achieve interrater reliability of 80 percent agreement on 95 percent of codes before coding the full set of interview data (Miles & Huberman, 1994). Once researchers achieve interrater reliability, the evaluation team will conduct an initial round of coding that focuses on categorizing the data into broad constructs. Each interview will be coded by one coder. The evaluation team will then engage in a second round of coding by applying inductive coding, during which patterns and emergent themes will be coded within each of the initial descriptive codes but also by participants and within schools. Throughout the analyses, the evaluation team will use concept mapping and memoing to explore, document, and verify emerging patterns in the experiences of teachers and school leaders.

d. Degree of Accuracy Needed

The evaluation team used the PowerUp! tool to calculate the number of schools required for the study (Dong & Maynard, 2013). The evaluation team estimated power for a fixed-effect blocked cluster random assignment design with the impact on level 1 outcomes (student or teacher 1) and treatment occurring at level 2 (school). The evaluation team calculated the minimum detectable effect size (MDES) with 80 percent probability using a two-tailed test, 0.05 level of significance, and 50 percent of schools assigned to treatment and control. The MDES for the student outcome was based on the following additional assumptions: an intraclass correlation (ICC) of 0.156 (Garet et al., 2011), level 1 covariates and level 2 covariates explaining 75 percent of the variability in outcome at their respective levels, and an average of 60 students per school. Prior studies showed that student- and school-level pretest measures can explain a considerable amount of variance at each level when examining student achievement outcomes (Bloom et al., 2007; Hedges & Hedberg, 2013; Westine et al., 2013). The MDES for teacher outcome was based on the following additional assumptions: an ICC of 0.20, a level 1 covariate explaining 50 percent of variability, and a level 2 covariate explaining 50 percent of the variability in outcome at their respective levels.

The anticipated sample sizes will provide an MDES of 0.46 SDs for teacher outcomes and an MDES of 0.19 SDs for student achievement outcomes. The evaluation team relied on meta-analyses of studies of empirical interventions to establish an effect size benchmark for student and teacher outcomes. A recent meta-analysis of 191 studies that are RCTs designed to improve the teaching or learning of math among U.S. preK–grade 12 students found an average effect size of 0.31 SDs on student math achievement, with effect sizes ranging from −0.60 to 1.23 SDs (Williams et al., 2022). Another meta-analysis of 95 experimental and quasi-experimental preK–12 STEM professional development and curriculum programs reported an average effect size of 0.21 SDs on student outcomes (Lynch et al., 2019). A meta-analysis by Hill et al. (2008) indicated that the average effect size on students' math achievement for middle school intervention studies was 0.27 SDs. For teacher outcomes, a meta-analysis of 60 studies of teacher coaching programs that employed causal research designs showed a pooled effect size of 0.49 SDs on teacher instructional practice outcomes (Kraft et al., 2018); another meta-analysis of 40 studies of randomized experiments of interventions directed at classroom practice found an average of 0.42 SDs based on classroom observations (Garrett et al., 2019). The estimated MDESs for the proposed

evaluation are generally consistent with the average effect sizes reported in these meta-analyses, indicating that the proposed evaluation is sufficiently powered to detect impacts on student and teacher outcomes that are of statistical and practical significance. However, the evaluation team expects that statistical power will be limited for the exploratory analyses (moderator analyses).

e. Unusual Problems Requiring Specialized Sampling Procedures

There are no unusual problems requiring specialized sampling procedures.

f. Use of Periodic (less frequently than annual) Data Collection to Reduce Burden

This project will collect data one time for recruitment and implementation. Teacher self-efficacy data will need to be collected more frequently than annually because the evaluation is occurring within one school year, and the measures will need to be assessed in September (baseline survey) and May (post survey) of the same school year. A longer period between data collection would make it difficult for the study team to meet the requirements for the efficacy study (by preventing baseline and follow-up data collection in the time frame necessary for the evaluation).

B3. Methods for Maximizing the Response Rate and to Deal With Nonresponse

The evaluation team is committed to obtaining complete data for this evaluation. Based on the evaluation team's prior experience with administering surveys to teachers in a variety of schools, districts, and states, the team expects the response rate for the teacher surveys to be at 85 percent for those individuals who have consented to participate in the study. The evaluation team will contact nonresponding teachers up to four times to encourage participation. Three follow-up email reminders will be sent to individual respondents in the event that responses are not obtained for the surveys (sample language for the initial and follow-up emails is provided in appendix C). The evaluation team will consider other modes of follow-up, including reminder letters and reminder phone calls if response rates are below expectation.

Although the evaluation team expects high response rates (90 percent) for the administrator implementation checklist (because schools volunteer for this study in order to receive the toolkit for free), nonresponse follow-up will be performed to ensure adequate response rates. The team anticipates a 100 percent response rate for teacher and leader interviews because interviewees will be selected from those individuals who are willing to participate.

In addition, several steps will be taken to maximize response rates. For example, sampled respondents will receive advance communications that explain the study, introduce REL Midwest, provide an assurance of confidentiality, and encourage them to participate to help refine the toolkit. Respondents also will be given a contact number to reach the evaluation team with questions. Finally, respondents will receive an incentive for participating in the study: $30 per teacher survey or teacher interview and $50 per school leader or facilitator interview.

The evaluation team anticipates a 100 percent response rate from Illinois districts on teacher and student administrative data. A key to achieving complete administrative data is tracking the data components from each district with e-mail and telephone contact to the appropriate parties to resolve issues of missing or delayed data files. All administrative data files will be reviewed for consistency and completeness. If a data file has too many missing values, the evaluation team will seek to obtain more complete responses by e-mail or phone.

If a key variable (outcome or covariate) has a response rate below 85 percent, the evaluation team will conduct a nonresponse bias analysis on that variable, following the National Center for Education Statistics Statistical Standards for surveys (see https://nces.ed.gov/statprog/2012/; Chapter 4). The nonresponse bias analysis will: (1) assess whether sample members with data and the original study sample differ on other observed characteristics by a substantial magnitude and (2) assess the most likely reasons for missing data.

B4. Test of Procedures

The evaluation will focus on measuring the toolkit's impact on three key outcomes: teacher self-efficacy for fraction computation and rate and ratio instruction, classroom practice for fraction computation and rate and ratio instruction, and students' abilities to solve fraction computation and rate and ratio problems.

Because teacher self-efficacy will be examined using existing reliable and validated measures (DePiper et al., 2019; McGee &Wang, 2014), the evaluation team does not plan to conduct additional testing of the measures. Instead, the evaluation team will conduct psychometric analysis to examine the reliability and construct validity of the measures, using the data obtained from the baseline survey for the evaluation, and will make any additional adjustments or refinements, if needed, for the post survey.

Teacher practices will be examined through classroom observations using an observation protocol adapted from the Middle School Mathematics Professional Development Impact Study sponsored by IES (Garet et al., 2010). To measure students' abilities to solve fraction computation and rate and ratio problems, the evaluation team has constructed a customized test by drawing on items from existing state standardized tests (released items or practice test). State assessment items have undergone analysis for validity and reliability, as well as review to remove bias, ensuring item functioning. (Note: OMB clearance for classroom observations and student assessment is not being sought. They are mentioned here as context and to provide a description of the full design of the study.)

The instruments and protocols to be used for the implementation measures (teacher survey, administrator checklist, and teacher and leader interviews) have been shared with AIR colleagues who were formerly employed as teachers or district administrators or colleagues with content expertise These critical colleagues reviewed the instruments for clarity, face validity of questions, and brevity. During their review, they also looked for: (1) whether the questions asked are clear, understandable and free of research jargon, and answerable; (2) whether the questions actually assess the intended constructs; and (3) whether the number and type of questions are appropriate (e.g., not redundant, focused enough to solicit clear

answers). These instruments will be further pilot tested in fall 2023, with fewer than nine respondents for each instrument.

B5. Names of Statistical and Methodological Consultants and Data Collectors

The following individual was consulted on the statistical aspects of the design:

Joshua Polanin, PhD, Principal Researcher, American Institutes for Research; (202) 403-5509; jpolanin@air.org

AIR, ED's contractor for REL Midwest, is conducting this project. Yinmei Wan is the principal investigator, and Melinda Griffin is the project director. The staff from REL Midwest contributing to the study methods, instrument development, and data collection are Rachel Garrett, Max Pardo, Kathryn Rich, Jingyan Xia, and Will Johnston.

References

Barbieri, C. A., Rodrigues, J., Dyson, N., & Jordan, N. C. (2020). Improving fraction understanding in sixth graders with mathematics difficulties: Effects of a number line approach combined with cognitive learning strategies. *Journal of Educational Psychology, 112*(3), 628–648. http://eric.ed.gov/?ID=EJ1247111

Bloom, H. S., Richburg-Hayes, L., & Black, A. R. (2007). Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis, 29*(1), 30–59. https://eric.ed.gov/?id=EJ782431

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology, 37*(4), 247–253. http://eric.ed.gov/?ID=EJ977998

DePiper, J. N., Nikula, J., & Louie, J. (2019). Shifts in self-efficacy for teaching English learners: Emergent findings from mathematics teacher professional development. In S. Otten, A. G. Candela, Z. de Araujo, C. Haines, & C. Munter (Eds.), *Proceedings of the forty-first annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 547–551). University of Missouri.

Dong, N., & Maynard, R. (2013). "PowerUp"!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness, 6*(1), 24–67. https://eric.ed.gov/?id=EJ994691

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

Garet, M. S., Wayne, A. J., Stancavage, F., Taylor, J., Eaton, M., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., Doolittle, F., & Warner, E. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE No. 2011–4024). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://eric.ed.gov/?id=ED519922

Garet, M. S., Wayne, A., Stancavage, F., Taylor, J., Walters, K., Song, M., Brown, S., Hurlburt, S., Zhu, P., Sepanik, S., & Doolittle, F. (2010). *Middle school mathematics professional development impact study: Findings after the first year of implementation* (NCEE No. 2010–4009). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://eric.ed.gov/?id=ED509306

Garrett, R., Citkowicz, M., & Williams, R. (2019). How responsive is a teacher's classroom practice to intervention? A meta-analysis of randomized field studies. *Review of Research in Education, 43*(1), 106–137. https://journals.sagepub.com/doi/10.3102/0091732X19830634

Harvey, R. (2012). Stretching student teachers' understanding of fractions. *Mathematics Education Research Journal, 24*, 493–511. https://eric.ed.gov/?id=EJ984997

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis, 29*(1), 60-87.

Hedges, L. V., & Hedberg, E. C. (2013). Intraclass correlations and covariate outcome correlations for planning two- and three-level cluster-randomized experiments in education. *Evaluation Review, 37*(6), 445–489. https://eric.ed.gov/?ID=EJ1034857

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*(3), 172–177.

Jones, D. L., Zientek, L. R., Sharon, V. V., & Swarthout, M. B. (2020). Solving equations with fractions: An analysis of prospective teachers' solution pathways and errors. School Science & Mathematics, 120(4), 232–243. https://doi.org/10.1111/ssm.12402Liu, Y. (2018). Fraction magnitude understanding and its unique role in predicting general mathematics achievement at two early stages of fraction instruction. British Journal of Educational Psychology, 88(3), 345–362. https://doi.org/10.1111/bjep.12182 Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547–588. https://doi.org/10.3102/0034654318759268

Liu, Y. (2018). Fraction magnitude understanding and its unique role in predicting general mathematics achievement at two early stages of fraction instruction. *British Journal of Educational Psychology, 88*(3), 345–362. https://doi.org/10.1111/bjep.12182

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base that informs stem instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis, 41*(3), 260–293. https://eric.ed.gov/?id=EJ1223474

McGee, J. R., & Wang, C. (2014). Validity-supporting evidence of the Self-efficacy for Teaching Mathematics Instrument. *Journal of Psychoeducational Assessment, 32*(5), 390–403. https://eric.ed.gov/?ID=EJ1030705

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Sage Publications.

National Center for Education Statistics (NCES). (2012). 2012 Revision of NCES Statistical Standards: Final. https://nces.ed.gov/statprog/2012

Park, J. H., Lee, I. H., & Cooc, N. (2019). The role of school-level mechanisms: How principal support, professional learning communities, collective responsibility, and group-level teacher expectations affect student achievement. *Educational Administration Quarterly, 55*(5), 742–780. https://eric.ed.gov/?id=EJ1232698

Pashley, N. E., & Miratrix, L. W. (2022). Block what you can, except when you shouldn't. *Journal of Educational and Behavioral Statistics, 47*(1), 69–100. https://eric.ed.gov/?ID=EJ1323829

Siegler, R., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., Thompson, L., & Wray, J. (2010). *Developing effective fractions instruction for kindergarten through 8th grade: A practice guide* (NCEE 2010-4039). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED512043

Siegler, S. F., & Lortie-Forgues, H. (2015). Conceptual knowledge of fraction arithmetic. *Journal of Educational Psychology, 107*(3), 909–918. https://doi.org/10.1037/edu0000025

Tekin-Sitrava, R. (2020). Middle school mathematics teachers' reasoning about students' nonstandard strategies: Division of fractions. *International Journal for Mathematics Teaching and Learning, 21*(1), 77–96.

Westine, C. D., Spybrook, J., & Taylor, J. T. (2013). An empirical investigation of variance design parameters for planning cluster-randomized trials of science achievement. *Evaluation Review, 37*(6), 490–519. https://pubmed.ncbi.nlm.nih.gov/24785938/

Whitehead, A. N., & Walkowiak, T. A. (2017). Preservice elementary teachers' understanding of operations for fraction multiplication and division. *International Journal for Mathematics Teaching & Learning, 18*(3), 293–317. https://eric.ed.gov/?id=EJ1164169

Williams, R., Citkowicz, M., Miller, D. I., Lindsay, J., & Walters, K. (2022). Heterogeneity in mathematics intervention effects: Evidence from a meta-analysis of 191 randomized experiments. *Journal of Research on Educational Effectiveness*. https://www.tandfonline.com/doi/full/10.1080/19345747.2021.2009072