AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

2022, Vol. 114, No. 3, 560–575
https://doi.org/10.1037/edu0000711

Promoting Understanding of Measurement and Statistical Investigation Among Second-Grade Students With Mathematics Difficulties

Christian T. Doabler[1], Ben Clarke[2], Derek Kosty[3], Marah Sutherland[2], Jessica E. Turtura[2], Allison R. Firestone[4], Georgia L. Kimmel[1], Patrick Brott[2], Tasia L. Brafford[1], Nancy J. Nelson Fien[5], Keith Smolkowski[3], and Kathleen Jungjohann[2]

[1] Meadows Center for Preventing Educational Risk, The University of Texas at Austin
[2] Center on Teaching and Learning, University of Oregon
[3] Oregon Research Institute, Eugene, Oregon, United States
[4] Graduate School of Education, University of California, Berkeley
[5] Wheelock College of Education and Human Development, Boston University

Measurement and statistical investigation are areas of mathematics visibly neglected in educational intervention research, particularly studies involving students with or at risk for mathematics difficulties (MD). This shortage is concerning given the importance these areas hold in students' pursuit of mathematical proficiency. This study investigated the initial efficacy of the *Precision Mathematics Grade 2* (PM-2) intervention, a Tier 2 (print and technology-based), integrated STEM intervention designed to increase second-grade students' mathematics achievement in the areas of measurement and statistical investigation. A total of 130 second-grade students with or at risk for MD participated in the randomized, controlled trial. Students were randomly assigned within classrooms to either treatment (PM-2) or control (business-as-usual) conditions. Findings indicated a pattern of "promise" for PM-2 improving student scores on a proximal assessment of early measurement skills (Hedges' g = .50). Differential effects of PM-2 by initial numeracy skill were not observed for 3 of the outcome measures. However, moderation results were found on a curriculum-based measure, suggesting the effects of PM-2 were greatest for students with higher initial skill. Implications for supporting students' development of mathematics proficiency in areas beyond whole numbers and operations are discussed.

Educational Impact and Implications Statement

A strong understanding of measurement and data is essential to becoming a contributing member of today's society. This study investigated whether a second-grade mathematics intervention improved the mathematics achievement of 130 second-grade students who were at risk for long-term difficulties in mathematics. The *Precision Mathematics Grade 2* intervention contains 32 lessons focused on foundational concepts of early measurement and data. Results indicated the intervention had positive effects on an early mathematics assessment. Students with higher mathematics skills at the start of the study were also found to receive the most benefit from the intervention. These findings suggest providing instruction in the areas of measurement and data may be important for supporting the mathematical learning of at-risk learners.

Keywords: efficacy, explicit mathematics instruction, mathematics difficulties, measurement, statistical investigation

Supplemental materials: https://doi.org/10.1037/edu0000711.supp

This article was published Online First December 30, 2021.
Christian T. Doabler https://orcid.org/0000-0002-8074-4311
Derek Kosty https://orcid.org/0000-0002-9792-3383
Marah Sutherland https://orcid.org/0000-0002-6108-8515
Allison R. Firestone https://orcid.org/0000-0003-0290-1788
Georgia L. Kimmel https://orcid.org/0000-0001-7649-3073
Nancy J. Nelson Fien https://orcid.org/0000-0002-9434-4888
Keith Smolkowski https://orcid.org/0000-0003-2565-3297

Research reported here was supported by the National Science Foundation through Grants 1503161 and 2010550 to The Meadows Center for Preventing Educational Risk at The University of Texas at Austin and the Center on Teaching and Learning at the University of Oregon.

Correspondence concerning this article should be addressed to Christian T. Doabler, Meadows Center for Preventing Educational Risk, The University of Texas at Austin, 1 University Station, D5300 SZB 408B, Austin, TX 78712, United States. Email: cdoabler@austin.utexas.edu

The National Council of Teachers of Mathematics (NCTM, 2000) holds a vision where the design and delivery of high-quality mathematics instruction affords all students equitable access to a coherent mathematics curriculum and a pathway for developing mathematical proficiency. Fundamental to achieving NCTM's priorities and goals is instruction focused on the areas of measurement and statistical investigation. Supporting students' mathematical competence in early measurement concepts and skills, for example, provides them a foundation for grasping how measurement tools function, understanding the attribute being measured, and making comparisons of the measured attribute relative to a particular unit (Goldenburg & Clements, 2014). Strong problem-solving skills around foundational concepts of measurement can also help lay the foundation for working with more advanced mathematics (Clements et al., 2018; Doabler, Clarke, Kosty, et al., 2019; Frye et al., 2013; Parmar et al., 2011; Sarama et al., 2021; Vasilyeva et al., 2020). For example, understanding that length is continuous and can always be partitioned into smaller lengths is key background knowledge for fractions (Beckmann, 2008). The NCTM also recognizes the importance of building a robust understanding of early statistical investigation (see Crites & St. Laurent, 2015). Organizing and representing continuous data on graphs, for instance, can lead to deeper knowledge of more sophisticated data analysis concepts, such as variability and sampling distributions (Langrall & Mooney, 2011; Litke & Hill, 2020; Shaughnessy, 2007).

In the early elementary grades (i.e., kindergarten to second grade), when students solve problems around measurement and early statistics, they have opportunities to apply and strengthen their understanding of whole numbers and operations. For example, when second-grade students solve computation problems embedded within a statistical investigation, they have opportunities to utilize and extend their conceptual and procedural knowledge of place value. Students can also apply their understanding of the underlying structures of word problem types (e.g., comparison problems) and engage in whole number computations when working with data-driven problems.

Combined, the areas of measurement and statistical investigation also serve as an optimal platform for teachers to take an integrated approach to science, technology, engineering, and mathematics (STEM) education (Honey et al., 2014; Moore et al., 2020). At the forefront of integrated STEM instruction are purposefully designed learning opportunities that center on real world situations and problems. These integrated opportunities are intended to support students' development of connected knowledge structures of fundamental ideas, concepts, and skills within and across STEM disciplines (Honey et al., 2014). In a second-grade classroom, a teacher might use a disciplinary core idea identified in the Next Generation Science Standards (NGSS, 2013) as an integrated context for deepening students' understanding of length measurement and data. For example, the teacher might design a set of lessons on how water can change the shape of coastal sand dunes. In these lessons, students would learn how to use different measurement tools to (a) investigate the effects of wave-induced erosion, (b) represent their measurements on a line plot, and (c) utilize the measurement data to answer questions about changes in dune height.

Given the advantages of situating foundational STEM literacy skills within measurement and statistical investigation it seems critical that students with or at risk for mathematics difficulties (MD) receive systematically designed opportunities to work with data and solve problems involving early concepts of measurement. Yet, measurement and statistical investigation are areas of mathematics often neglected in educational intervention research. The shortage of methodologically rigorous studies involving these areas is particularly worrisome given their importance in supporting students' development of mathematical proficiency. As such, this study sought to contribute to the current knowledge base by investigating the initial efficacy of a Tier 2 second-grade mathematics intervention designed to increase at-risk students' understanding of measurement and statistical investigation.

Status of Measurement and Statistical Investigation in U.S. Mathematics

Never before has knowledge of measurement and data been so critical to becoming an informed citizen. In current times, for example, it is imperative to read and understand political polls, make important decisions about findings from clinical drug trials, and engage in public policy debates (e.g., global warming). Well-designed instruction focused on foundational concepts of measurement and statistical investigation also provides the opportunities that students require to develop advanced skills needed for future positions in STEM (Clements et al., 2020). A rich understanding of statistical investigation, for example, can assist epidemiologists in understanding the epidemic spread of a virus, such as COVID-19 (Fauci et al., 2020).

Despite the importance of promoting students' mathematical proficiency with measurement and data analysis, compelling evidence suggests a considerable number of U.S. students, particularly students from subgroups at risk for MD, struggle to reach proficient levels in these areas of mathematics (National Center for Education Statistics [NCES], 2019). Results of the 2019 fourth-grade National Assessment of Educational Progress (NAEP) revealed that students eligible for free or reduced lunch programs, Hispanic students, and multilingual students scored .7 to 1.0 standard deviations lower than their peers on the Measurement subscale and the Data Analysis, Statistics, and Probability subscale.

Such data present challenges to schools for how to accelerate the learning of U.S. students who struggle with the foundational concepts and skills of measurement and statistical investigation. Although a host of different factors likely influence the learning difficulties these students face, here, we consider two plausible ones. First, when considering the enacted curriculum in today's mathematics classrooms, little instruction occurs on building students' quantitative literacy, such as knowledge of data and statistics. For example, when examining direct observations and teacher transcripts obtained from 144 upper-elementary classrooms (i.e., fourth and fifth grade), Litke and Hill (2020) found scarce instruction on data analysis. Of the data-focused instruction observed, most centered on graph construction and using calculations to work with data. Another key finding of this study was that classrooms spent a nominal amount of instructional time focused on the scientific practices associated with statistical investigation, such as formulating important research questions, planning and carrying out such investigations, and constructing explanations to interpret findings.

Although measurement is perhaps a more common instructional focus in the elementary grades, measurement instruction often targets procedural skills such as measuring object lengths, instead of building a deeper conceptual understanding of foundational measurement principles (Clements et al., 2018; National Research Council, 2009). For example, Polikoff (2012) found that use of measurement instruments was one of the three most redundant mathematics topics addressed across kindergarten and first grade classrooms, covered by 89% and 91% of classroom instruction, respectively. Importantly, early measurement instruction should first focus on the underlying concepts of measurement, with care taken in the transition between informal understanding of measurement principles and formal measurement skills such as using measuring tools (Beckmann, 2008; Irwin et al., 2004; Sutherland et al., 2020).

In combination with the lack of focused instruction on measurement and statistical investigation is the dearth of educational intervention research on these mathematical areas. National calls have pushed for increasing students' understanding of number systems (National Mathematics Advisory Panel, 2008) and accordingly a wealth of research focused on rigorously testing whole number interventions has emerged (e.g., Dyson et al., 2013). Unfortunately, this has been at the neglect of other critical areas of mathematics, including measurement and statistical investigation. Reviews conducted by nationally recognized independent evaluators confirm this lack of a scientific imprint. Of the 18 elementary mathematics intervention programs reviewed by the National Center on Intensive Intervention (NCII, n.d.), not one focused specifically on measurement and statistical investigation. Similar findings have been reported by the What Works Clearinghouse (WWC, n.d.).

In sum, the absence of focused core instruction and paucity of intervention research underscore the need for researchers to iteratively design and rigorously test interventions focused on the critical concepts and skills of early measurement and statistical investigation. By developing and establishing such interventions, we contend the field will be better positioned to support the gestalt of mathematical proficiency among at-risk learners.

Iterative Design of a Second-Grade Mathematics Intervention

Recognizing the need for increasing student mathematics achievement in the areas of measurement and statistical investigation, our research team developed the *Precision Mathematics Grade 2* (PM-2) program. PM-2 is a 32-lesson, Tier 2 intervention designed for second-grade students with or at risk for MD and intended for delivery in small group formats (i.e., four to five students per interventionist), 30-min per day, four days per week for eight weeks. Conceptually, PM-2 is designed to directly supplement core (Tier 1) mathematics instruction. In this way, students receive PM-2 in addition to their core mathematics instruction.

Because the current landscape largely ignores the call from instructional designers to embrace a scientific footprint in the development of mathematics programs (Carnine et al., 1997; Zucker et al., 2019), our research team used the Curriculum Research Framework (CRF) proposed by Clements (2007) during our PM-2 intervention development efforts. The CRF comprises three categories of development activities: a priori foundations, learning models, and evaluation. Embedded within the three categories, Clements proposes 10 phases of the curriculum development and evaluation process. While the initial phases of the CRF focus on researchers collecting early implementation data on intervention prototypes, the later phases encourage researchers to scale up their investigations to establish the impact of intervention programs in more summative research.

In our work with the PM-2 program, we applied the initial phases of the CRF (Clements, 2007) in a series of iterative cycles of development, implementation, field testing, analysis, and review. These cycles offered opportunities to collect initial empirical data about PM-2 implementation and eventually establish the intervention's feasibility, usability, and accessibility prior to the current study. For example, earlier research ensured the intervention could be implemented in the timeframe allotted for instruction (i.e., 30-minute sessions), utilized by end-users as designed with few implementation bottlenecks (e.g., hang or freeze of student tablets), and had contextual fit in authentic educational settings (Doabler, Clarke, Firestone, et al., 2019).

At the forefront of the PM-2 intervention is an explicit instructional design platform (Coyne et al., 2011) that directly teaches the critical concepts and problem-solving skills of measurement and statistical investigation identified in the Common Core State Standards for Mathematics for second grade (CCSS-M, 2010), including, among others, partitioning, length-unit iteration, indirect comparisons, alignment of the zero-point on a ruler, and categorial data. Our team elected to target the areas of measurement and statistical investigation for several reasons. First, establishing an early understanding of measurement and statistics is essential for setting a pathway to becoming mathematically literate in today's society. Second, a strong body of empirical evidence suggests that as students exit the lower elementary grades, many face deficits in these areas (NCES, 2019). Third, as argued in the Progressions for the Common Core Standards in Mathematics (Common Core Standards Writing Team [CCSWT], 2011), the areas of measurement and statistical investigation complement each other well and together represent important an important backdrop to support students' development of mathematical proficiency. Additionally, when students engage in the fundamental concepts and skills of measurement and statistical investigation it reflects the work of professionals in the STEM fields. For example, to investigate important questions focused on solutions to pressing problems in school systems, researchers measure observable behaviors and use data analysis techniques to interpret their findings. As such, the PM-2 intervention was purposefully designed to build and strengthen students' conceptual understanding of these understudied areas of mathematics by facilitating frequent, meaningful student practice opportunities around technology-based activities and hands-on investigations.

The PM-2 intervention offers daily technology-based activities intended to build students' conceptual understanding and problem-solving skills. Our rationale for engineering these types of activities within the intervention largely stems from the push to integrate education technology into today's classrooms (Atkins et al., 2010). Although the empirical literature on education technology is still in its infancy (WWC, n.d.), well-designed technology can make students' learning experiences more accessible and individualized (Mayer, 2009). Moreover, technology can spark students' interest, pique their curiosity, and help motivate their learning

(Atkins et al., 2010). As such, in PM-2 students receive their own iPads and use them through scaffolded support from the teacher and her iPad. For example, one lesson has students work with a virtual measuring tool to measure and compare the height of sunflowers. Students then record their virtual measurements on a bar graph and utilize the data to solve a set of "compare" word problems to determine the height differences between the sunflowers.

Complementing the technology-based activities are hands-on investigations intended to extend students' understanding of targeted mathematical content and further their pursuit of overall mathematical proficiency. Such investigations, often designed within the same lessons as the technology-based activities, promote tactile experiences for students to work with concrete objects and different measurement tools. For example, experts suggest that students build conceptual understanding when they make connections among concepts and procedures in different situations and in different ways (Kilpatrick et al., 2001). Therefore, the hands-on investigations offer opportunities for students to solve new problems and to verbalize their understanding through small-group mathematics discourse opportunities.

To promote connections between and within STEM disciplines (Honey et al., 2014; Moore et al., 2020), the PM-2 intervention comprises eight "Investigative Units" that are grounded in the Disciplinary Core Ideas within the second-grade Life Science and Earth Science topics in the Next Generation Science Standards (NGSS, 2013). Units are designed to be delivered across four consecutive days, for 30 minutes per lesson. The NGSS were integrated into the curriculum to provide students with meaningful, real-world contexts to solve early measurement and data analysis problems, while providing the added benefit of increased exposure to important scientific ideas and phenomena. For example, to foster deeper learning of "take-away" mathematics word problems, the intervention situates these problem types within the context of how wind and water can change the shape of the land (NGSS, 2013).

Purpose of the Study

The purpose of this study was to investigate the initial efficacy of the PM-2 intervention for improving the mathematics achievement of second-grade students with or at risk for MD. PM-2's development was funded through a federally-funded design and development project (Doabler, Clarke, Fien, Nelson, et al., 2015). Prominent funding agencies, such as the Institute of Education Sciences (IES) and the National Science Foundation (NSF), expect such development projects to demonstrate "promise" of an intervention for leading to desired student outcomes (The Institute of Education Sciences, U.S. Department of Education and the National Science Foundation, 2013). Generating sufficient initial efficacy evidence can warrant conducting future efficacy trials to more formally test the treatment effects of a newly developed intervention. As such, the current research examines results from a recent pilot study of PM-2 that employed a randomized controlled trial design and included 130 second-grade students with or at risk for MD.

In addition to determining the intervention's initial efficacy, there was also strong interest in testing whether students derive differential response from the PM-2 intervention. Recent research suggests students' initial skill is a risk factor for early MD and a plausible explanatory variable of response variation among students who receive early mathematics interventions (Clarke et al., 2019; Fuchs et al., 2019; Powell et al., 2017; Vasilyeva et al., 2020). In line with those studies, we examined whether and to what extent students' whole number skills prior to the start of the PM-2 intervention served as a student-level predictor of differential response to the PM-2 intervention. Prior understandings of whole numbers and operations may influence the extent to which a student responds to instruction focused on measurement and statistical investigation. Consequently, an insufficient understanding of whole numbers and operations may hamper a student's capacity to understand and solve problems in these more complex mathematical contexts. However, we hypothesized initial mathematics skill would not moderate the treatment effects of the PM-2 intervention. The rationale for this hypothesis was based on the systematic and explicit nature of the PM-2 intervention. One of the most consistent findings of mathematics intervention research is that systematic and explicit mathematics instruction is effective for the full range of at-risk learners (Fuchs et al., 2021; Gersten et al., 2009). Therefore, we expected PM-2 to benefit the entire treated sample and that the effects would be of similar magnitude across the range of preintervention skills in whole numbers at the start of the intervention. Two research questions were addressed:

1. Is there *initial evidence of efficacy* of the PM-2 intervention on the mathematics achievement of second grade students at risk for MD?

2. Is there *initial evidence of differential response* to the PM-2 intervention as a function of pretest performance on a distal measure of whole number understanding?

Method

This study analyzed efficacy data collected during a federally-funded design and development project (Doabler, Clarke, Fien, Nelson, et al., 2015). The PM-2 study employed a randomized controlled trial and was conducted during the 2018–2019 school year in a school district located in the Northwest region of the U.S. Second-grade students were randomly assigned within classrooms to one of two conditions: (a) the PM-2 intervention or (b) a control (i.e., business-as-usual) condition. We elected to use this rigorous research design to establish the initial efficacy of the PM-2 intervention and better position it for more formal efficacy testing in future research (The Institute of Education Sciences, U.S. Department of Education and the National Science Foundation, 2013).

Participants

Schools

Classified as being situated in a large suburban locale, the participating school district contained four elementary schools, all of which participated in the current study. The district had an enrollment of 5,338 students. Within the four schools, less than 1% of students identified as American Indian or Native Alaskan; less than 2% to 4%, Asian; less than 1%, Black; 5% to 21%, Hispanic; less than 1%, Native Hawaiian or Pacific Islander; 69% to 85%, White; and 5% to 8%, more than one race. Within these same schools, 11% to 16% of students received special education services; less than 5% to 10% were multilingual students; and 8% to

34% were eligible for free or reduced-price lunch. These student demographics were representative of other large suburban districts in the state in which the current research was conducted.

Classrooms

To recruit participating classrooms, our research team initially contacted the district's curriculum director and mathematics coordinator, who then consulted with the principals and mathematics coaches at each elementary school. Following these conversations, we then recruited all 13 of the district's second-grade teachers. Thus, 13 second-grade teachers or classrooms participated in the study. All classrooms provided mathematics instruction in English and operated five days per week. The thirteen classrooms each had an average of 25.5 students (SD = 2.5). Of the 13 teachers, all identified as female. Teachers had an average of 18.6 (SD = 8.1) years of teaching experience and 10.7 (SD = 8.5) years of teaching in second grade classrooms. Ten of the teachers had a master's degree in education, and eight had completed an algebra course at the college level.

Students and Inclusion Criteria

As approved by our institutional review board, this study used a passive parental permission process that included sending a passive consent notification letter to parents of all students in the 13 participating classrooms. Our notification letter to parents described the study's purpose, its duration and desired outcomes, and procedures to maintain confidentiality of student data. Parents then had the opportunity to decline their child's participation by returning a preaddressed, prepared postcard. Of the 330 notification letters mailed, three (1%) were returned and thus opted out of the study. Then, in each participating classroom, all students with parental consent were screened in the winter of their second-grade year. The screening process, which included 327 second-grade students, comprised the four measures of the Texas Early Mathematics Inventory-Progress Monitoring (TEMI-PM; University of Texas System/Texas Education Agency, 2008): Magnitude Comparisons, Number Sequences, Place Value and Addition/Subtraction Combinations. In each participating classroom, an independent evaluator rank ordered students' TEMI-PM Composite scores and selected the lowest 10 students for study inclusion. Four alternate students (the next lowest ranking students) were also preselected in each classroom.

Prior to random assignment, each participating teacher reviewed their classroom's list of the 10 lowest-ranking students and removed any students with chronic absenteeism, significant behavioral issues, or conflicts with receiving their special education instruction. In cases where teachers removed a student from the list, the student was replaced by one of the preselected alternates. Application of the alternate-student model occurred with 18 students, of which 15 were removed based on scheduling conflicts with ongoing special education services and three owing to poor attendance.

The independent evaluator then created pairs of adjacent students with respect to their TEMI-PM Composite scores. One student from each pair was then randomly assigned to the PM-2 intervention (treatment) and the other to the control condition (business-as-usual). Of the 327 students who were screened, 130 were randomly assigned within each of the 13 classrooms to the treatment (n = 65) or the control conditions (n = 65). It is important to note that to reduce the threat of compensatory rivalry our research team did not reveal to control students the outcomes of the random assignment process. In all, 13 intervention groups were formed (one per classroom), each with five students.

Demographic data indicated that 17% of treatment and 11% of control students received special education services, 6% and 5% were multilingual students, 25% and 15% received free or reduced lunch, and 63% and 57% were female, respectively. Although the majority racial group for both conditions identified as White (75%), 16% of treatment and 11% of control students were Hispanic, and 11% and 2% were Multiple Races, respectively. The average student age at pretest was 7.7 years (SD = .4) in the treatment and control conditions.

Interventionists

PM-2 intervention groups were taught by 13 interventionists, of which 12 were district-employed instructional assistants and one was hired specifically for this study. Each interventionist taught one PM-2 intervention group. Among the interventionists, all identified as female. Most had previous experience with small group mathematics instruction (53%) and held a bachelor's degree or higher (67%). Interventionists had an average of 6.7 (SD = 4.8) years of teaching experience and 53% had taken an algebra course at the college level.

Procedure

PM-2 Intervention

PM-2 is a Tier 2 mathematics intervention designed to build students' conceptual understanding and problem-solving skills in the area of early measurement and statistical investigation. The scope and sequence of the intervention draws from three of the four second-grade standards within the measurement and data domain of the CCSS-M (2010), including (a) measure and estimate lengths in standard units, (b) relate addition and subtraction to length, and (c) represent and interpret data. A preponderance of empirical evidence suggests that explicit and systematic instruction is the most effective method for teaching mathematics to students at-risk for MD (Fuchs et al., 2021; Gersten et al., 2009). Therefore, the program was engineered to embrace an explicit and systematic instructional design framework. Lessons are fully scripted with activities structured to include overt teacher modeling, scaffolded opportunities for students to work with concrete and representational mathematical models and engage in mathematics verbalizations, and academic feedback to confirm correct answers and to address errors and potential misconceptions.

To support students' conceptual understanding of statistical investigation, the PM-2 intervention utilizes a framework of statistical investigation issued by the American Statistical Association (Franklin et al., 2007). The framework, outlined in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report, includes four components: (a) formulate research questions; (b) collect relevant data; (c) organize and represent data on graphs, and analyze data to address the targeted research question; and (d) interpret the results. Students learn why the components are important and how to apply them in different measurement and data-driven problem-solving contexts. For example, an activity situated in the context of ecosystems teaches students the importance of making precise observations of plant growth after exposure to sun and water. Students are then afforded opportunities to make sense of their measurements and justify their conclusions.

In the area of measurement, the intervention prioritizes several big ideas of length measurement. These concepts and skills include (a) partitioning, (b) length-unit iteration, (c) transitivity, (d) alignment of the zero-point on a ruler, and (e) the inverse relationship between the size of the unit of measure and the number of those units (CCSWT, 2011; Clements & Stephan, 2004). Additionally, students learn to solve addition and subtraction word problems involving lengths. To build comprehension of these targeted concepts and procedures, PM-2 offers frequent opportunities for students to represent them in different science-related situations. For example, PM-2 introduces the concept of partitioning, which refers to the understanding that the length of an object can be subdivided into smaller equal sized units, in the context of students using centimeters and inches to measure the length of different animals. In a subsequent set of lessons, the intervention provides explicit opportunities for students to make connections with the concept of partitioning by having them use virtual yard sticks to measure the above sea level height of icebergs.

Along with developing students' mathematical proficiency with early concepts of measurement and statistical investigation, the PM-2 intervention prioritizes two other foundational aspects of early mathematics. The first is clear and concise mathematical language (Fuchs et al., 2021; Riccomini et al., 2015). Students have the opportunity to practice using precise mathematical language specific to measurement and data analysis (e.g., *centimeter, height, line plot, pictograph*) along with academic vocabulary that is used more broadly across content areas (e.g., *explain, investigation, combine*). Given the rich literature base behind direct vocabulary instruction (e.g., Coyne et al., 2010), the PM-2 intervention utilizes this instructional approach by explicitly introducing the terms in student-friendly language, and offering systematic opportunities for review and practice incorporated across the Investigative Units. These opportunities are important for multilingual students and other students who demonstrate academic risk because they allow them to (a) hear and practice the correct pronunciation of key mathematical terms, (b) understand the meaning of targeted words in multiple contexts, and (c) verbally apply their understanding of the target words by using them in complete sentences.

The second aspect prioritized in PM-2 is the word problem subtypes specified in the CCSS-M (2010) for second grade. Specifically, PM-2 lessons address add-to, take-from, put-together, take-apart, and compare situations, with the unknowns or missing information mostly in the third position of equations (e.g., $8 - 5 = x$). Students are explicitly taught how to use a schema-broadening instructional approach (Powell, 2011) to identify and use the structural features underlying each problem type to guide their solution. Following the introduction of new problem types, word problems are integrated into the lessons to allow for practice embedded in the context of measurement and statistical investigation.

PM-2 relies upon a "hybrid" framework (Doabler, Clarke, Firestone, et al., 2019) that integrates technology with print-based activities and hands-on problem-solving investigations to promote a robust understanding of critical measurement and data analysis concepts and problem-solving skills among students with MD. Measurement activities, such as physically iterating a unit to measure an object's length, or working with formal measurement tools such as rulers or meter sticks, are introduced through hands-on and technology-based activities designed to promote conceptual understanding and generalization of measurement concepts. The technology-based activities offered on the PM-2 app are operated on iPads. All iPad-based activities are controlled by a teacher iPad, with students following along on Wi-Fi–connected individual iPads to allow for individualized instruction and practice. The teacher iPad offers a "teacher view" component, which displays student data boxes across the top of the screen to show how students respond to real-time questions on their respective iPads. Across the tech-and print-based mediums of instruction, students are able to engage in mathematical discourse with their teacher and peers around core concepts of measurement and data analysis.

The following provides an example of a lesson that incorporates both hands-on and iPad-based activities (see sample lesson for Week 5, Day 2 in the online supplemental materials). To begin, the teacher leads students through vocabulary practice on the iPads to review key words and their definitions, such as inch, length, meter, and the word problem types. Special attention is given to the "compare" word problem type, which is used later on in the lesson during the statistical investigation. Students then use standard rulers to measure three paths that different bees took when pollinating flowers, shown on a worksheet. In this activity, the teacher introduces the concept of the zero point (i.e., the end of the ruler) and the distance to the first hash mark (1 in.), and has students use inch tiles to confirm the length-units of the ruler. Students record the lengths of the paths in inches on the worksheet. The teacher then leads students through a statistical investigation on their iPads to answer the question, "How many more bees traveled 30 in. than 40 in.?" Students use virtual rulers on the iPad to measure the lengths of various bee paths and record distances on a line plot. They are last guided to use their graphical data to solve the compare word problem and to write a corresponding equation.

In this study, the PM-2 intervention was delivered in 30-min, small group formats (i.e., five students per interventionist), four days per week for approximately eight weeks, beginning in early winter. Because PM-2 is designed as a supplemental intervention, treatment students continued to receive core mathematics instruction during the study. It is also important to note that treatment students did not receive Tier 2 mathematics intervention support above and beyond the PM-2 intervention. Moreover, to further reduce the threat of compensatory rivalry and possibility of resentful demoralization among control students (Shadish et al., 2002), the PM-2 intervention was delivered in settings outside the general education classroom.

Professional Development

All interventionists participated in two 4-hr professional development workshops delivered by project staff. The first workshop was held prior to the start of PM-2 implementation and focused on the instructional objectives and content of Lessons 1–16, small group management strategies (e.g., managing group-level mathematics discourse), and effective instruction using iPads. The second workshop targeted new objectives and content introduced in Lessons 17–32. Across workshops, interventionists were provided opportunities to practice and receive feedback on lesson delivery from project staff. To promote implementation fidelity, all interventionists received approximately two in-class coaching visits from members of the curriculum development team, who had specialized knowledge and training in evidence-based mathematics instruction and the PM-2 intervention. Coaching centered around implementation of the PM-2 intervention and effective small group instructional practices and included feedback on instructional quality and fidelity of intervention implementation. Results indicated coaches' postobservation debriefings with the interventionists lasted an average duration of 6.9 minutes (*SD* = 3.8).

Fidelity of Implementation

Fidelity of implementation was measured via direct observations by members of the curriculum development team. Data were aggregated to the interventionist level to describe average ratings observers gave each interventionist across visits (mean number of visits = 1.8, range = 1–3). Observers documented 19 features of intervention implementation, including presence of 16 core principles of instruction and 3 impressions of instructional quality. Descriptive statistics for ratings of the presence of core principles (coded 1 'not at all' to 5 'fully present') and overall impressions (coded 1 = *low* to 5 = *high*) are presented in Table 1. Results indicated that all core principles were almost fully present and overall impressions were high. To mitigate the potential for contamination or treatment diffusion (Shadish et al., 2002), access to PM-2's technology-based activities, which represent a core component of the intervention, was managed through a password protection process. Therefore, participating interventionists were required to enter their unique identification number on the teacher and student iPads for each lesson. Data on entry to the intervention was then collected via our institute's database. In monitoring this information, we found no evidence to suspect treatment contamination (e.g., excessive intervention usage on a daily basis).

Table 1
Descriptive Statistics for Fidelity of Implementation of PM-2

Measure	*M* (*SD*)
Core principles of instruction	
Prepare for lesson	4.8 (0.4)
Organization of materials	5.0 (0.1)
Seating conducive to task	4.8 (0.3)
Students engaged in lesson	4.3 (0.6)
Students follow routines and expectations	4.4 (0.6)
Tasks modeled for students	4.7 (0.4)
Guided practice provided	4.7 (0.3)
Individual practice provided	4.7 (0.3)
Clear group response signal used	3.9 (0.7)
Feedback is specific and timely	4.4 (0.6)
Affirmation of correct responses	4.1 (0.8)
Immediate correction of errors	4.3 (0.5)
Additional practice provided	4.5 (0.4)
Objectives met with high success	4.4 (0.5)
Pace of instruction appropriate	4.4 (0.6)
Use of iPads appropriate	4.7 (0.4)
Overall impressions of instructional quality	
Fidelity of implementation	4.8 (0.4)
Quality of student–teacher interactions	4.3 (0.5)
Quality of interaction with technology	4.7 (0.4)

Note. PM-2 = Precision Mathematics Grade 2. Fidelity data were based on coaching visits to all interventionists. Ratings of the presence of core principles were coded 1 = *not at all* to 5 = *fully present*, and overall impressions were coded 1 = *low* to 5 = *high*.

Control Condition

The control condition, which consisted of school-based Tier 1 and Tier 2 instructional supports, was documented through teacher surveys. For their core (Tier 1) mathematics instruction, all teachers reported using Bridges in Mathematics, a PK–5 mathematics curriculum that is aligned with the CCSS-M and delivered in a variety of instructional formats, including whole-class, small-group, and 1:1 settings. Teachers noted that they provided an average of 59 min (*SD* = 21.58) of daily core mathematics instruction. Most teachers (82%) indicated that teaching number and operations in base ten was prioritized in their core instruction, followed by operations in algebraic thinking. As reported by teachers, approximately 11 of the 65 (17%) control students received Tier 2 mathematics intervention support in addition to their core instruction. The majority of teachers reported using the Bridges in Mathematics intervention program during the Tier 2 instruction. Such supports, which began in September and ran until the remainder of the school year, were delivered in small-group formats, four days per week in 30-min sessions, which was an equivalent dose of Tier 2 instruction to that of PM-2 during the study. The reported focus of Tier 2 control instruction was on finding patterns in numbers, fact families, identifying the relationship between addition and subtraction, and decomposing and composing numbers.

Mathematics Outcome Measures

Four mathematics outcome measures were administered at pretest and posttest. These measures included a proximal assessment that measured skills identified in the CCSS-M (2010) for second grade and three distal mathematics outcome measures. Trained research staff administered the measures and interscorer reliability criteria ≥ .95 were met for all assessments.

Precision Mathematics-2 Proximal Assessment is an individually administered measure that comprises 24 items related to topics identified under the measurement and data domain in the CCSS-M (2010) for second grade. In an untimed setting, students are asked to measure objects, identify appropriate units for objects and the relationship between measurement units, create and interpret graphs, and identify word problem types. Coefficient alpha for the current study sample was .58 at pretest and .68 at posttest. Concurrent validity correlations with the Early Measurement-Curriculum Based Measure and Texas Early Mathematics Inventory-Progress Monitoring scores were .21 and .25 at pretest and .29 and .46 at posttest, respectively.

Texas Early Mathematics Inventory-Progress Monitoring (TEMI-PM; University of Texas System/Texas Education Agency, 2008) assesses students' number sense and understanding of quantity. For second grade, the TEMI-PM consists of four individually administered two-minute measures including (a) Magnitude Comparisons, (b) Number Sequences, (c) Place Value, and (d) Addition/Subtraction Combinations. Alternate form reliability of the TEMI-PM is reported at greater than .75 for all subtests (Bryant et al., 2008). Concurrent validity with the SAT-10 and ITBS ranges from .50 to .72 and predictive validity with the SAT-10 ranges from .57 to .65 (Bryant et al., 2008). In the current study, we used the total score generated from all four TEMI-PM measures (i.e., TEMI-PM Composite) as one of our primary distal outcome measures of student mathematics achievement. In our sample, coefficient alpha for the composite score was .98 at pretest and posttest and the concurrent validity correlation with the Early Measurement-Curriculum Based Measure was .53 at pretest and .51 at posttest.

Word Problems (Jordan & Hanich, 2000) is an individually administered, assessment of students' word problem solving skills. Students are asked to solve 14 word problems focused on money (i.e., pennies) and ranged from conceptually simple to conceptually complex. Four types of story problems are assessed: change problems, combine problems, compare problems, and equalize problems, with the missing variable in all positions (e.g., addends, minuends, total). All problems involve basic number combinations with sums and minuends up to 9 in an untimed format. Coefficient alpha for the current study sample was .76 at pretest and .83 at posttest. Concurrent validity correlations with the Early Measurement-Curriculum-Based Measures (EM-CBM) Composite and TEMI-PM scores in the current sample were .31 and .52 at pretest and .37 and .50 at posttest, respectively.

EM-CBM (Clarke et al., 2021) is an individually administered set of four, researcher-developed, one-minute fluency-based CBMs focused on the broader concepts of early measurement identified in the CCSS-M (2010) for second grade. The EM-CBM battery comprises four distal measures: (a) Comparison of Three Objects, (b) Iteration-1, (c) Iteration-2, and (d) Measurement of Two Items Using an Object. These measures assess students' understanding of length comparisons of multiple objects, length comparison of two objects using a third, and iterative measurement using a nonstandard measurement tool. EM-CBM is a timed assessment, therefore, coefficient alphas were not calculated (Anastasi, 1988; Burns & Dobson, 1981). The correlation between pretest and posttest EM-CBM scores was .66. The concurrent validity correlation with the TEMI-PM Composite was .53 at pretest and .51 at posttest.

Statistical Analysis

We tested the initial efficacy of the PM-2 intervention using a mixed Time × Condition analysis designed to account for students (unit of random assignment) nested within small groups in the treatment condition and not nested in the control condition (Baldwin et al., 2011; Bauer et al., 2008). These partially nested analyses tested for differences between conditions on changes in outcomes from pretest to posttest and are described in detail by Doabler et al. (2016). The model included effects of condition, time, and the Time × Condition interaction, with condition coded 0 for control and 1 for treatment and time coded 0 at pretest and 1 at posttest. The effect of condition represents the difference in outcome between the treatment and control conditions at pretest. The effect of time represents the change in outcome from pretest to posttest among the control condition. The Time × Condition interaction represents the difference in change in outcome between the treatment and control conditions. We also examined whether pretest TEMI-PM scores predicted differential response to the PM-2 intervention by expanding the statistical model to include the moderator and its interaction with condition, time, and the Time × Condition term. The three-way interactions provided estimates of whether condition effects varied by pretest mathematics skill levels.

Model Estimation

We performed the analyses with SAS PROC MIXED Version 9.4 (SAS Institute, 2016), restricted maximum likelihood estimation, and Satterthwaite approximation to determine degrees of freedom. Maximum likelihood estimation uses all available data and produces potentially unbiased results even in the face of substantial missing data, provided the missing data were missing at random (Schafer & Graham, 2002). We considered this assumption tenable given the minimal rates of missing data. The statistical model also assumes independent and normally distributed observations. We addressed the first of these assumptions by modeling the multilevel nature of the data. The outcome measures in the present study also did not markedly deviate from normality; skewness and kurtosis fell within ±2.3.

Effect Sizes and Interpretation of Results

Hedges' (1981) g values were calculated to characterize the magnitude of treatment effects (WWC, 2020). In response to the recommendations of the American Statistical Association (Wasserstein & Lazar, 2016), we abstained from using bright-line rules, such as from claims of "statistical significance" when $p < .05$. p values have an interpretation as a measure of incompatibility between the observed data and all assumptions of the statistical model including the null hypothesis, H_0 (Greenland et al., 2016; Wasserstein & Lazar, 2016). This cumbersome definition neither informs on which assumptions are incorrect nor the importance of the association. To complement p values and Hedges' g values, we report Akaike weights (Akaike, 1973), which describe the strength of evidence for one model when comparing it with others. Akaike weights—also called model probabilities—express the probability of a model given a set of competing models and the observed data (Burnham et al., 2011) and can be interpreted as the probability that the same model would be selected with a "replicate data set from the same system" (Burnham et al., 2011, p. 30). They quantify the strength of evidence for each hypothesis, represented by a statistical model, given the data and all other hypotheses (models) tested. For each analysis, we compared models for two hypotheses, one with the intervention effect (H_A) and one without (H_0), and reported the Akaike weight, w, for the model with the condition effect (H_A). With only two models, the model probability for H_0, the model without the condition effect, is $1 - w$. Hence, $w = .75$ suggests that the probability of H_A is .75, whereas the probability of H_0 is .25. This roughly corresponds to a result with $p = .05$, which implies the limited value of "just-significant" results ($p \approx .05$). In such a case, the model for H_A has an approximately 75% chance of being the best-fitting model; in other words, the model for H_A is only three times as likely as the model for H_0 given the data. We considered model probabilities at or above .75 or $p < .05$ as statistical evidence of an effect.

Results

Baseline Equivalence and Attrition

Table 2 presents means, standard deviations, and sample sizes for the four outcome measures by assessment time and condition. Treatment and control groups did not differ on pretest outcomes

Table 2
Descriptive Statistics for Mathematics Outcome Measures by Assessment Time and Condition

Measure	Pretest		Posttest	
	Intervention	Control	Intervention	Control
PM-2 proximal assessment				
M (*SD*)	12.9 (3.1)	13.0 (3.2)	17.8 (3.0)	16.3 (3.6)
n	63	64	64	63
EM-CBM composite				
M (*SD*)	46.0 (13.5)	45.9 (15.7)	61.3 (14.1)	58.6 (16.2)
n	59	56	61	59
TEMI-PM composite				
M (*SD*)	82.1 (15.1)	81.1 (17.4)	96.7 (18.4)	96.1 (22.8)
n	65	65	64	63
Word problems				
M (*SD*)	10.2 (3.0)	10.3 (3.1)	10.9 (3.1)	11.4 (3.1)
n	64	63	64	63

Note. EM-CBM = Early Measurement-Curriculum-Based Measures; PM-2 = Precision Mathematics Grade 2; TEMI-PM = Texas Early Mathematics Inventory-Progress Monitoring; The sample included 65 treatment students and 65 control students. EM-CBM = Early Measurement – Curriculum Based Measures; TEMI-PM = Texas Early Mathematics Inventory-Progress Monitoring.

(gs $\leq$ .07, ps $\geq$ .6640). Examination of attrition between pretest and posttest revealed 98% (64/65) of the treatment and control students completed a posttest assessment.

Evidence of Intervention Effects for PM-2

We first tested the hypothesis that students in the intervention condition experienced greater increases in mathematics outcomes than did students in the control condition. Among those students identified as eligible for PM-2, we found differential gains favoring students in the intervention condition on the PM-2 proximal assessment (g = .50 [.08, .91], t_{19} = 2.51, p = .0210; w = .84). The effect size was medium, and the model estimate for differences in gains between conditions was 1.6. Limited evidence of intervention effects was observed with the EM-CBM (g = .18 [−.15, .51], t_{48} = 1.11, p = .2710; w = .39), TEMI-PM (g = −.01 [−.23, .21], t_{34} = −.09, p = .9268; w = .26), and Word Problems assessment (g = −.13 [−.35, .10], t_{72} = −1.13, p = .2613; w = .39). See Table 3 for complete model results.

The model probabilities in the first row of Table 3 suggest that the hypothesis of condition effects was likely to describe the data for PM-2 proximal assessment; the model with condition effects had a considerably higher weight (w) of .84 compared with the model without condition (1 – w = .16). The model probabilities suggest that the hypothesis of condition effects was unlikely to describe the data for the other three outcomes; the models without condition effects had considerably higher weights of (1 – w) .61, .74, and .61 for the EM-CBM, TEMI-PM, and Word Problems assessment, respectively. Likelihood ratio tests summarized in the bottom two rows of Table 3 indicated that homoscedastic residuals fit better than heteroscedastic residuals for all student outcomes. Time × Condition effects were similar in the heteroscedastic and homoscedastic models.

Evidence of Differential Response

Next, we examined whether pretest TEMI-PM scores moderated the PM-2 intervention effects. To do so, we added this variable and its interaction with condition, time, and the Time × condition term to the model. Table 4 presents tests of differential response to PM-2 as a function of pretest TEMI-PM scores for each respective student mathematics outcome. Tests of the three-way interactions provided evidence that pretest TEMI-PM scores moderated intervention effects on EM-CBM scores (t_{112} = 2.75, p = .0070; w = .91). For these models, w describes the model probability for the model with the test of moderation compared with an equivalent model without the necessary three-way interaction. Therefore, the EM-CBM model with the moderation effect (i.e., three-way interaction) was more likely than the model without (w = .91). Differential effects of PM-2 by preintervention numeracy skill levels were not observed for the PM-2 Proximal and Word Problems assessments.

Figure 1 illustrates the differential effects of PM-2 on the EM-CBM based on pretest TEMI-PM scores. The graph depicts the estimated difference between conditions (vertical axis) across the range of pretest TEMI-PM scores (horizontal axis) surrounded by confidence bounds. Zero on the vertical axis represents no difference between conditions. The vertical lines within the graph represent sample percentiles, not unlike a boxplot. The vertical lines show that about 25% of the students received pretest TEMI-PM scores below 75.0 and 5% below 56.0. The confidence bounds exclude zero at TEMI-PM scores of 90.0 or higher, or for students scoring above the 68th percentile (42nd percentile within the TEMI-PM norm-referenced sample). This suggests that only the 32% of the students with the highest pretest TEMI-PM scores demonstrated differential gains in EM-CBM scores favoring the PM-2 intervention condition.

Discussion

This article reports findings from a pilot study aimed at testing the initial efficacy of PM-2, a recently developed Tier 2 intervention designed to teach at-risk second-grade students the foundational concepts and skills of early measurement and statistical investigation identified in the CCSS-M (2010). A total of 130 second-grade students with or at risk for MD participated in the randomized controlled trial, with students randomly assigned within classrooms to treatment (PM-2) or control (BAU) conditions. Two research questions were addressed.

Table 3
Results of a Partially Nested Time × Condition Analysis of Pretest to Posttest Change in Mathematics Outcomes

Effect or Statistic	PM-2 Proximal	EM-CBM	TEMI-PM	WP
Model probability (*w*)	.84	.39	.26	.39
Fixed effects				
Intercept	13.0*** (0.4)	45.5*** (1.9)	81.1*** (2.3)	10.2*** (0.4)
Time	3.3*** (0.4)	12.8*** (1.9)	14.8*** (1.6)	1.2*** (0.3)
Condition	−0.1 (0.7)	0.0 (2.7)	1.1 (4.2)	0.0 (0.6)
Time × Condition	1.6* (0.7)	2.8 (2.5)	−0.2 (2.3)	−0.4 (0.3)
Variances				
Group-level intercept	1.6 (1.2)	−0.1 (18.9)	108.6 (58.3)	1.5 (1.1)
Group-level gains	0.7 (0.9)	−4.7 (7.8)	2.7 (8.5)	−0.4* (0.2)
Student-level pre–post covariance	3.3** (1.2)	126.5*** (31.1)	166.1*** (43.2)	5.5*** (1.2)
Residual	4.7*** (1.0)	99.0*** (15.9)	74.4*** (12.2)	3.0*** (0.4)
Hedges' g [95% CI]				
Time × Condition	0.50 [0.08, 0.91]	0.18 [−0.15, 0.51]	−0.01 [−0.23, 0.21]	−0.13 [−0.35, 0.10]
p value				
Time × Condition	.0,210	.2,710	.9,268	.2,613
BH *p* value				
Time × Condition	.0,840	.3,613	.9,268	.3,613
df				
Time × Condition	19	48	34	72
Likelihood ratio χ^2	3.45	2.25	2.77	4.30
P value	.1,778	.3,248	.2,500	.1,163

Note. EM-CBM = Early Measurement-Curriculum-Based Measures; PM-2 = Precision Mathematics Grade 2; TEMI-PM = Texas Early Mathematics Inventory-Progress Monitoring; WP = Word Problems. Table entries show parameter estimates with standard errors in parentheses except for model probabilities (Akaike weights), Hedges' *g* values, *p* values, Satterthwaite degrees of freedom (*df*), and χ^2 values. The model probabilities indicate the likelihood of the model that contains the condition effect compared with the model without the condition effect given the data. *p* values provided with and without the Benjamini-Hochberg (BH) correction. Likelihood ratio tests compared homoscedastic residuals to heteroscedastic residuals with a criterion α of .10 and two degrees of freedom.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Initial Intervention Efficacy

Our first research question examined the intervention's proximal effects on a researcher-developed assessment related to topics identified under the measurement and data domain in the CCSS-M. Distal effects were investigated with a word problem-solving assessment and a norm-referenced, general outcome measure focused on number, operations, and quantitative reasoning. A set of four CBMs focused on foundational concepts of early measurement were also administered to examine distal outcomes.

Overall, findings were mixed. We found preliminary evidence of PM-2 effects on the PM-2 proximal assessment, with treatment students demonstrating greater gains than their control peers across the study period. The effect size (Hedges' *g*) was medium, and the Akaike weight (*w*) implied that the model with the condition effect was more likely than the model that assumed the null hypothesis of no condition effect (.84 vs. .16). A small positive effect size (g = .18) was also estimated for the measurement CBMs, but the Akaike weight implied that the model with the condition effect was slightly less likely than a model that assumed the null hypothesis (.39 vs. .61). We found no statistical evidence of positive intervention effects on the Word Problems assessment or the broader measure that assessed students' foundational number and numeration skills (TEMI-PM). Given the targeted focus of the PM-2 intervention, the lack of transfer to a set of distal measures primarily comprising number concepts perhaps should not be surprising. Nonetheless, it is plausible that the short duration of the PM-2 intervention (32 lessons) affected the lack of generalization of skills to the TEMI-PM and Word Problems assessments. The skills taught in the PM-2 intervention may require more time to develop and eventually transfer to these more distal outcomes. Thus, administration of a delayed posttest might provide time for the skills learned in PM-2 to marinate and impact other skills (e.g., whole number computation).

Encouragingly, preliminary evidence of "promise" emerged for improving students' proximal outcomes on a researcher-developed measure. Although the measure aligns with the instructional goals of the PM-2 intervention, its items were purposefully developed to directly reflect the measurement and data analysis topics identified in the CCSS-M. Therefore, we contend that it taps into the broader mathematical concepts and skills that students are expected to learn in second grade. Given that measurement and data analysis compose critical areas of mathematics and students' mathematical learning, these findings, although preliminary, add beginning support for the field to extend assessment and intervention work in areas of mathematics beyond whole numbers and early number sense skills (Vasilyeva et al., 2020).

Our findings of larger effects on a more proximal-aligned measure and no significant effects on broader outcome measures are similar to that commonly reported in mathematics (see Gersten, 2016) and reading intervention studies (e.g., Pressley et al., 2006; Swanson et al., 2017). Thus, it raises questions in the present study for how targeted mathematics interventions, such as PM-2, can help students transfer their knowledge to other contexts and outcomes. For example, word problem instruction is a main fixture of the PM-2 intervention. Yet, nonsignificant negative effects were found on the Word Problems assessment (Jordan & Hanich, 2000), a measure focused on total, difference, and change word problem types. Even though the Word Problems measure is contextualized around money (i.e., pennies), we expected treatment

Table 4
Tests of TEMI-PM Scores as Predictor of Differential Response to Intervention

Effect or Statistic	PM-2 Proximal	EM-CBM	TEMI-PM	WP
Model probability (*w*)	.34	.91	.61	.36
Fixed effects				
Intercept	13.0*** (0.4)	44.7*** (1.8)	81.6*** (1.1)	10.2*** (0.3)
Time	3.3*** (0.4)	13.5*** (1.8)	14.8*** (1.5)	1.2*** (0.3)
Condition	−0.2 (0.6)	0.7 (2.3)	0.0 (1.6)	−0.1 (0.4)
Time × Condition	1.6* (0.6)	1.5 (2.5)	−0.1 (2.3)	−0.4 (0.4)
TEMI	0.0 (0.0)	0.5*** (0.1)	1.0*** (0.1)	0.1*** (0.0)
TEMI × Condition	0.0 (0.0)	−0.3 (0.2)	0.0 (0.1)	0.0 (0.0)
TEMI × Time	0.1** (0.0)	−0.2* (0.1)	0.1 (0.1)	0.0 (0.0)
TEMI × Time × Condition	0.0 (0.0)	0.5** (0.2)	−0.1 (0.1)	0.0 (0.0)
Variances				
Group-level intercept	−0.1 (0.9)	−18.3 (9.9)	0.0 (5.9)	−0.4 (0.4)
Group-level gains	0.8 (0.9)	−0.9 (8.9)	3.1 (8.4)	−0.2 (0.1)
Student-level pre–post covariance	3.6** (1.1)	120.8*** (26.1)	0.0 (8.4)	
Residual	4.3*** (0.9)	87.7*** (15.0)	72.5*** (11.9)	
PM-L1 residual				2.0*** (0.4)
Pre–post covariance				5.1*** (1.3)
Control Residual				3.5*** (0.6)
Pre–post covariance				3.7*** (1.0)
P value				
TEMI × Time × Condition	.3,496	.0,070	.2,749	.2,934
BH *p* value				
TEMI × Time × Condition	.3,496	.0,280	.3,496	.3,496
df				
TEMI × Time × Condition	124	112	116	98
Likelihood ratio χ^2	2.86	2.42	1.20	5.11
P value	.2,390	.2,978	.5,493	.0,779

Note. EM-CBM = Early Measurement-Curriculum-Based Measures; PM-2 = Precision Mathematics Grade 2; TEMI-PM = Texas Early Mathematics Inventory-Progress Monitoring; WP = Word Problems. Table entries show parameter estimates with standard errors in parentheses except for model probabilities (Akaike weights), *p* values, Satterthwaite degrees of freedom (*df*), and χ^2 values. The model probabilities indicate the likelihood of the model that contains the condition effect compared with the model without the condition effect given the data. *p* values provided with and without the Benjamini-Hochberg (BH) correction. Likelihood ratio tests compared homoscedastic residuals to heteroscedastic residuals with a criterion α of .10 and two degrees of freedom.
$^{*} p < .05$. $^{**} p < .01$. $^{***} p < .001$.

students to generalize the problem-solving skills acquired during the PM-2 intervention to this measure. Perhaps the lack of impact on the Word Problems assessment was attributable to the fact that the measure contains items with the missing variable in all positions (e.g., addends, minuends, total), whereas the PM-2 intervention primarily targets problems with missing information in the more common third position of equations (e.g., $8 - 5 = x$). Regardless of the reason, future revisions to the PM-2 intervention may be necessary to best help students become stronger problem solvers.

Differential Response

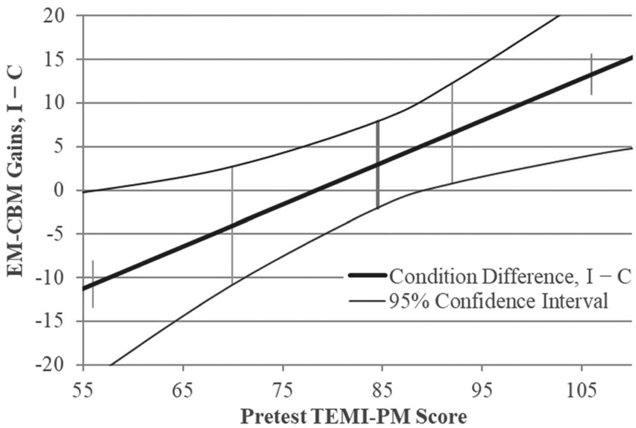
Our second research question examined for differential response to the PM-2 intervention. Although efficacy data are an integral component for building the knowledge base on effective instructional practices and establishing evidence behind academic interventions, alone these data are insufficient to grasp why an intervention works, for whom, and under what conditions. As such, the field has begun to further the call for educational researchers to investigate students' preexisting knowledge of foundational academic content as a predictor of students' responsiveness to such treatments (Fuchs & Fuchs, 2019). However, recent explorations into initial skill status and intervention response have shown incongruous results. For example, some studies have found students with lower initial skills benefit most from academic interventions (Clarke et al., 2019), whereas others have shown differential response benefiting students with higher pretest achievement (Vaughn et al., 2019). Conversely, findings from other research point to no moderation effects by initial skill levels (Fuchs et al., 2019).

Based on the instructional design of PM-2, we anticipated positive intervention effects across the range of at-risk students and little evidence of differential response owing to initial skill. However, for the battery of experimental CBMs, only students with the highest pretest TEMI-PM scores benefited from the intervention. This mixed evidence of differential response as a function of preintervention mathematics achievement suggests that future edits to PM-2 may be necessary to better ensure its effects are comparable regardless for where students' mathematics skill levels fall on a continuum of at-risk performance.

Limitations

In light of interpreting the current study's findings and planning for future intervention design and development work in the areas of measurement and statistical investigation, several limitations should be considered. One limitation was our small sample size and underpowered statistical analyses. Nonetheless, the sample size included in the current study is comparable to other pilot studies where the unit of randomization is at the student level (e.g., Bryant et al., 2011; Dyson et al., 2013). On a related note, our

Figure 1
Differential Effects of PM-2 on the EM-CBM Scores Based on Pretest TEMI-PM Scores



Note. EM-CBM = Early Measurement-Curriculum-Based Measures; PM-2 = Precision Mathematics Grade 2; TEMI-PM = Texas Early Mathematics Inventory-Progress Monitoring. The vertical axis shows the difference between conditions—zero on the vertical axis represents no difference between conditions—and the horizontal axis represents the range of pretest TEMI-PM scores. The heavy increasing line depicts the mean difference between conditions at each pretest value. The two thinner, outer lines show the 95% confidence interval around the mean estimate. To show the location of the sample on the graphs, the vertical lines within each figure depict the median (heavier vertical line), 25th and 75th percentiles (thinner long lines), and the 5th and 95th percentiles (short outer lines). For example, a TEMI-PM score of about 70.0 represents the lower 25th sample percentile at pretest.

sample mostly comprised students who identified as White. Although student demographics in the participating district were representative of other large suburban districts in the state in which the current research was conducted, working with a more diverse group of students would increase the generalizability of our findings.

Another limitation was our method for using the pretest performances of the 10 lowest students per classroom to identify our student sample. Based on some of our previous mathematics intervention work in early elementary classrooms (Fien et al., 2016), we anticipated that selecting the 10 lowest students of a class of 25 students would result in a study sample of at-risk students. Although our approach in the current study was less rigorous than using a defined cutoff score, it should be noted that the process employed resulted in an average TEMI-PM pretest score for our analytic sample that corresponded to the 27th percentile and that 67% of the student sample scored below the 40th percentile. Recent syntheses suggest that student performance at or below the 40th percentile is considered at-risk for MD (Nelson & McMaster, 2019; Nelson & Powell, 2018). Therefore, results from the present study may generalize to students who score slightly above the TEMI-PM threshold for supplemental intervention eligibility. Regardless, to improve the rigor of our screening process, future research should employ a defined cut score of initial mathematics skill.

Relatedly, our inclusion process allowed for teachers to remove students prior to random assignment. However, for the vast majority of students, their exclusion was based on exogenous factors beyond our control. For example, 15 of the 18 excluded students had scheduling conflicts between their special education services and the daily delivery of the PM-2 intervention. Despite this, it is worth noting that the percentage of students who remained in our sample and received special education services still equaled the percentages reported at the participating school levels.

The use of a passive control (business-as-usual) group also hampered our ability to conclude that PM-2 was more effective than a particular intervention. Prior to conducting the current study, we made an a priori decision for incrementally building evidence of “promise” around the newly designed PM-2 intervention. Thus, a passive control condition was a most logical first step in our intervention development and testing process. Relatedly, while information on the nature of mathematics instruction delivered in the control condition was documented through teacher surveys, a more nuanced measurement tactic, such as direct observation, may have offered a more vivid picture of the counterfactual, a deeper explanation of the current study’s findings, and a more robust method of monitoring treatment contamination. For example, results indicated nonsignificant negative effects on the word problem measure. A direct observation system may have revealed that control students received more frequent opportunities to practice solving different word problems than their PM-2 peers.

Furthermore, there were limitations around assessment measures owing to available resources and funding. For example, one major challenge encountered in the current study was the lack of available assessments to capture students’ skills in the domain of measurement and statistical investigation. Given this paucity of outcome measures, we believe our colleagues who conduct research in the areas of measurement and statistical investigation have encountered similar shortages. To combat this shortage, our team developed two of the study’s outcome measures, the PM-2 Proximal Assessment and the EM-CBM measures. It is important to note that despite being researcher developed, neither measure was tailored specifically to the PM-2 intervention. Rather each was developed to map directly onto the broader big ideas of length measurement and statistical investigation identified in the second-grade CCSS-M (2010). In this way, treatment and control students had equal opportunity to access and learn the mathematical content purported to be assessed by these two outcome measures, such as length-unit iteration and direct comparisons. Although we carefully developed both assessments using an iterative development and user testing process (Clements, 2007; Doabler, Clarke, Fien, Baker, et al., 2015), the current findings indicate the PM-2 proximal assessment had relatively low reliability and validity evidence. Consequently, revisions to this measure are necessary to improve its psychometric properties and permit caution free interpretation of intervention effects.

Finally, owing to funding constraints, we were unable to include follow-up assessments to explore for fadeout and persistence of the observed treatment effects on the proximal assessment and broader measurement CBMs. Because a common finding of educational research is the disappearance of intervention effects in a posttreatment period (Kang et al., 2019), the administration of follow-up measures in design and development projects could help inform intervention revisions to promote stronger persistence of treatment impact.

Implications for Research and Practice

Despite these limitations, the current study has a number of implications for the field, including demonstrating preliminary promise of a Tier 2 mathematics intervention focused on building students' mathematical proficiency in the areas of measurement and statistical investigation. Compared with similarly skilled students in the control condition, students receiving the PM-2 intervention demonstrated greater gains on a proximal outcome measure that targeted topics from the measurement and data domain in the CCSS-M (2010). It is encouraging that gains were observed on this measure given the focus of the PM-2 intervention and the importance of foundational measurement and statistical investigation skills to accessing more advanced mathematical concepts, such as the rational number system and algebraic reasoning (NMAP, 2008). Recognizing that students at risk for MD disproportionately struggle with concepts of measurement and statistical investigation (NCES, 2019), the findings of the current study provide initial support for accelerating students' knowledge in these areas of mathematics through a relatively low dose of a supplemental Tier 2 intervention (i.e., 32 lessons).

Also worthy of discussion is the lack of robust statistical evidence for treatment effects on the EM-CBM, TEMI-PM, or Word Problems assessments, two of which focus on a broader range of mathematical skill including whole number understanding and word problem solving, respectively. Although PM-2 embeds opportunities for students to use and expand upon their whole number and problem-solving skills within the contexts of measurement and statistical investigation, it is possible that second graders struggling with mathematics would benefit from more targeted support in the area of whole number and operations specifically prior to receiving the PM-2 intervention. Theoretically, this multi-intervention approach would prime students' background knowledge and skills in whole numbers and thus allow them to become more successful when engaging in interventions focused on "other" areas of mathematics. As such, we encourage the field to investigate whether there is added value to using a multiintervention approach to accelerate the mathematics outcomes of students with MD. Relatedly, the PM-2 intervention uniquely embraces an integrated STEM approach (Moore et al., 2020), allowing students to apply their mathematical knowledge in rich science-related contexts. While science outcomes were not assessed in the current study, we encourage the field to investigate STEM integration given its potential to help at-risk learners transfer their knowledge across STEM disciplines.

Taking our findings into consideration, schools looking to adopt new mathematics intervention programs should consider multiple factors, including the specific areas of need for students with MD, the provision of school resources such as availability of staff to implement multiple intervention programs (e.g., teaching both whole number interventions and those focused on topics beyond whole numbers), and the alignment of interventions with content coverage in core instruction. Unfortunately, measurement and statistical investigation are typically minimized in elementary, core mathematics instruction (Litke & Hill, 2020), which raises the question of how an intervention focused on measurement and statistical investigation topics would align with Tier 1 mathematics instruction. We argue that the strong links between understanding of foundational measurement principles and later critical mathematics concepts (e.g., the principle of iteration supporting student understanding of unit fractions), for example, necessitate a stronger focus to be placed on the development of measurement skills early on within core mathematics instruction and as a focus of intervention. Similar long-term implications are assumed for students developing early conceptual understanding of data and statistical investigation. Particularly for students with or at risk for MD, measurement and statistical investigation represent practically useful areas to develop mathematical proficiency as well as engaging and relevant platforms to practice other mathematical skills such as whole number operations and problem-solving skills.

We would also be remiss to not mention the important role that technology played in the PM-2 intervention, allowing for greater individualization and a wider range of instructional examples. The hybrid format of PM-2 allowed students to physically work with concrete models and measurement tools, such as centimeter cubes, rulers, and yardsticks, and to extend that learning by using virtual rulers and measurement tools on the iPad. Online virtual manipulatives are becoming increasingly available and more commonly used in classrooms. When used appropriately, virtual manipulatives provide some advantages over physical manipulatives, such as increasing the number and types of problems students encounter (Satsangi et al., 2016). In PM-2, the iPad-based component minimized the number of physical materials needed during instruction while simultaneously allowing for varied and engaging problem-solving contexts.

Technology was also leveraged to support interventionists in effectively implementing PM-2. For example, the display of student response data via the "teacher view" on the teacher iPad allowed interventionists to see whether students were correctly answering problems. These unique and innovative teacher vantage points allowed for real-time instructional differentiation, such as providing confirmations, error corrections, and additional student practice opportunities. Moreover, while students having their own iPads allowed for more frequent and individualized practice, interventionists could still control the pacing of student iPads through the teacher iPad, keeping the entire group on task. Given the advantages that technology can contribute to making mathematical learning opportunities more accessible and engaging for at-risk learners, and potentially more manageable for teachers, we would advocate for curriculum developers to explore embedding education technology within future mathematics intervention programs.

Finally, although the value of researcher-developed assessments is often more heavily questioned than commercially available ones, like our colleagues in the field of reading research (Clemens & Fuchs, 2021), we contend researcher-developed assessments can play an important role in testing for forms of learning transfer. Researcher-developed outcome measures are also instrumental for better understanding why academic interventions work, for whom, and under what conditions. For example, findings generated by the EM-CBM, a more distal based assessment, indicated evidence of differential response, such that only students with the highest pretest TEMI-PM scores benefited from the intervention. It may be that students with lower preintervention numeracy skills were unable to transfer what was taught in the PM-2 intervention to an outcome measure of generalized performance. Another plausible explanation is that students with higher preintervention numeracy skills were better prepared than their lower skilled peers to solve problems focused on length-unit iteration and direct comparisons in a timed setting. However, the intervention groups in the current study were heterogeneously formed, containing the full range of at-risk learners. Therefore, future tailoring of the PM-2 intervention may be necessary to address the entire range of TEMI-PM preintervention skill profiles.

Conclusion

Little research has been conducted on the areas of measurement and statistical investigation, particularly studies involving students with or at risk for MD. What is more, recent research suggests these areas of mathematics are rarely part of the enacted curriculum in elementary mathematics classrooms. These absences are especially concerning given the academic and practical implications for understanding the foundational concepts of measurement and possessing the problem-solving skills for investigating categorical and continuous data. The current study sought to address these blank spots in the literature by testing the initial efficacy of a Tier 2 mathematics intervention focused on concepts and problem-solving skills around measurement and statistical investigation. Although preliminary, results from a proximal outcome measure indicated a pattern of promise. However, transfer to the study's broader measures of mathematics was not detected. Furthermore, moderation results suggested that treatment effects were largest for students with higher preintervention numeracy skills. Our plan is to use these data and the lessons learned from the current pilot study to extend the program of research on the PM-2 intervention by investigating its efficacy in an adequately powered efficacy trial. We anticipate the findings obtained from this more rigorous line of research will allow us to provide the field with a feasible, empirically validated mathematics intervention focused on two important areas of early mathematics.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrox & F. Caski (Eds.), *Second international symposium on information theory* (pp. 267–281). Akademiai Kiado.

Anastasi, A. (1988). *Psychological testing*. McMillan Publishing Company.

Atkins, D. E., Bennett, J., Brown, J. S., Chopra, A., Dede, C., Fishman, B., Gomez, L., Honey, M., Kafai, Y., Luftglass, M., Pea, R., Pellegrino, J., Rose, D., Thille, C., & Williams, B. (2010). Transforming American education: Learning powered by technology. http://www.ed.gov/sites/default/files/netp2010.pdf

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

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. https://doi.org/10.1080/00273170802034810

Beckmann, S. (2008). *Mathematics for elementary teachers*. Pearson Addison Wesley.

Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children*, *78*(1), 7–23. https://doi.org/10.1177/001440291107800101

Bryant, B. R., Bryant, D. P., Shih, M., Sperling, R., Kim, S. A., & Seo, Y.-J. (2008). *Technical manual for the Texas early mathematics-inventories: Progress monitoring and Texas early mathematics inventories: Outcome grades K, 1, and 2*. The University of Texas System/Texas Education Agency.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23–35. https://doi.org/10.1007/s00265-010-1029-6

Burns, R. B., & Dobson, C. B. (1981). Reliability and validity. In *Experimental Psychology*. Springer. https://doi.org/10.1007/978-94-011-7241-7_25

Carnine, D., Jitendra, A. K., & Silbert, J. (1997). A descriptive analysis of mathematics curricular materials from a pedagogical perspective: A case study of fractions. *Remedial and Special Education*, *18*(2), 66–81. https://doi.org/10.1177/074193259701800201

Clarke, B., Doabler, C. T., Smolkowski, K., Turtura, J., Kosty, D., Kurtz-Nelson, E., Fien, H., & Baker, S. K. (2019). Exploring the relationship between initial mathematics skill and a kindergarten mathematics intervention. *Exceptional Children*, *85*(2), 129–146. https://doi.org/10.1177/0014402918799503

Clarke, B., Sutherland, M., Doabler, C. T., Kelsey, N., & Landis, B. (2021). Developing and investigating the promise of early measurement screeners. *School Psychology Review*. Advance online publication. https://doi.org/10.1080/2372966X.2021.1919493

Clemens, N. H., & Fuchs, D. (2021). Commercially developed tests of reading comprehension: Gold standard or fool's gold? *Reading Research Quarterly*. Advance online publication. https://doi.org/10.1002/rrq.415

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education*, *38*, 35–70. https://doi.org/10.2307/30034927

Clements, D. H., & Stephan, M. (2004). Measurement in pre-K to grade 2 mathematics. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 299–320). Erlbaum.

Clements, D. H., Sarama, J., Van Dine, D. W., Barrett, J. E., Cullen, C. J., Hudyma, A., Dolgin, R., Cullen, A. L., & Eames, C. L. (2018). Evaluation of three interventions teaching area measurement as spatial structuring to young children. *The Journal of Mathematical Behavior*, *50*, 23–41. https://doi.org/10.1016/j.jmathb.2017.12.004

Clements, D. H., Vinh, M., Lim, C., & Sarama, J. (2020). STEM for inclusive excellence and equity. *Early Education and Development*. Advance online publication. https://doi.org/10.1080/10409289.2020.1755776

Common Core State Standards Initiative. (2010). *Common core standards for mathematics*. http://www.corestandards.org/the-standards/mathematics

Common Core State Standards Writing Team. (2011). *Common core state standards for mathematics*. https://commoncoretools.files.wordpress.com/2011/06/ccss_progression_md_k5_2011_06_20.pdf

Coyne, M. D., Kame'enui, E. J., & Carnine, D. (2011). *Effective teaching strategies that accommodate diverse learners* (4th ed.). Pearson Education.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli, R., Jr., Ruby, M., Crevecoeur, Y. C., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness*, *3*(2), 93–120. https://doi.org/10.1080/19345741003592410

Crites, T., & St. Laurent, R. (2015). *Putting essential understanding of statistics into practice in grades 9-12*. National Council of Teachers of Mathematics.

Doabler, C. T., Clarke, B., Fien, H., Baker, S. K., Kosty, D. B., & Strand Cary, M. (2015). The science behind curriculum development and evaluation: Taking a design science approach in the production of a tier 2 mathematics curriculum. *Learning Disability Quarterly*, *38*(2), 97–111. https://doi.org/10.1177/0731948713520555

Doabler, C. T., Clarke, B., Fien, H., Nelson, N. J., & Smolkowski, K. (2015). *Precision Mathematics: Using interactive gaming technology to* *build student proficiency in the foundational concepts and problem solving skills of measurement and data analysis*. National Science Foundation, Division of Research on Learning, Directorate for Education and Human Resources, Research on Learning in Formal and Informal Setting.

Doabler, C. T., Clarke, B., Firestone, A. R., Turtura, J. E., Jungjohann, K. J., Brafford, T. L., Sutherland, M., Nelson, N. J., & Fien, H. (2019). Applying the curriculum research framework in the design and development of a technology-based tier 2 mathematics intervention. *Journal of Special Education Technology*, *34*(3), 176–189. https://doi.org/10.1177/0162643418812051

Doabler, C. T., Clarke, B., Kosty, D. B., Kurtz-Nelson, E., Fien, F., Smolkowski, K., & Baker, S. K. (2016). Testing the efficacy of a tier 2 mathematics intervention: A conceptual replication study. *Exceptional Children*, *83*(1), 92–110. https://doi.org/10.1177/0014402916660084

Doabler, C. T., Clarke, B., Kosty, D., Turtura, J. E., Firestone, A. R., Smolkowski, K., Jungjohann, K., Brafford, T. L., Nelson, N. J., Sutherland, M., Fien, H., & Maddox, S. A. (2019). Efficacy of a first-grade mathematics intervention focused on measurement and data analysis. *Exceptional Children*, *86*(1), 77–94. https://doi.org/10.1177/0014402919857993

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities*, *46*(2), 166–181. https://doi.org/10.1177/0022219411410233

Fauci, A. S., Lane, H. C., & Redfield, R. R. (2020). Covid-19—Navigating the uncharted. *The New England Journal of Medicine*, *382*(13), 1268–1269. https://doi.org/10.1056/NEJMe2002387

Fien, H., Doabler, C. T., Nelson, N. J., Kosty, D. B., Clarke, B., & Baker, S. K. (2016). An examination of the promise of the NumberShire level 1 gaming intervention for improving student mathematics outcomes. *Journal of Research on Educational Effectiveness*, *9*(4), 635–661. https://doi.org/10.1080/19345747.2015.1119229

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report*. American Statistical Association.

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide*. National Center for Education and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Fuchs, D., & Fuchs, L. S. (2019). On the importance of moderator analysis in intervention research: An introduction to the special issue. *Exceptional Children*, *85*(2), 126–128. https://doi.org/10.1177/0014402918811924

Fuchs, L. S., Fuchs, D., & Gilbert, J. K. (2019). Does the severity of students' pre-intervention math deficits affect responsiveness to generally effective first-grade intervention? *Exceptional Children*, *85*(2), 147–162. https://doi.org/10.1177/0014402918782628

Fuchs, L. S., Newman-Gonchar, R., Schumacher, R., Dougherty, B., Bucka, N., Karp, K. S., Woodward, J., Clarke, B., Jordan, N. C., Gersten, R., Jayanthi, M., Keating, B., & Morgan, S. (2021). *Assisting students struggling with mathematics: Intervention in the elementary grades* (WWC 2021006). National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. http://whatworks.ed.gov/

Gersten, R. (2016). What we are learning about mathematics interventions and conducting research on mathematics interventions. *Journal of Research on Educational Effectiveness*, *9*(4), 684–688. https://doi.org/10.1080/19345747.2016.1212631

Gersten, R. M., Chard, D., 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. https://doi.org/10.3102/0034654309334431

Goldenburg, E. P., & Clements, D. H. (2014). *Developing essential understanding of geometry and measurement for teaching mathematics in prekindergarten grade 2*. National Council of Teachers of Mathematics.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. https://doi.org/10.1007/s10654-016-0149-3

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Honey, M., Pearson, G., & Schweingruber, H. A. (Eds.). (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research* (Vol. 500). National Academies Press.

Irwin, K. C., Vistro-Yu, C. P., & Ell, F. R. (2004). Understanding linear measurement: A comparison of Filipino and New Zealand children. *Mathematics Education Research Journal*, *16*(2), 3–24. https://doi.org/10.1007/BF03217393

Jordan, N. C., & Hanich, L. B. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities*, *33*(6), 567–578. https://doi.org/10.1177/002221940003300605

Kang, C. Y., Duncan, G. J., Clements, D. H., Sarama, J., & Bailey, D. H. (2019). The roles of transfer of learning and forgetting in the persistence and fadeout of early childhood mathematics interventions. *Journal of Educational Psychology*, *111*(4), 590–603. https://doi.org/10.1037/edu0000297

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics*. Mathematics Learning Study Committee, National Research Council.

Langrall, C. W., & Mooney, E. S. (2011). Data analysis and probability. In F. M. Fennell (Ed.), *Achieving fluency: Special education and mathematics* (pp. 219–242). National Council of Teachers of Mathematics.

Litke, E., & Hill, H. C. (2020). Exploring Central Tendencies: The teaching of data and statistics in elementary mathematics classrooms. *The Elementary School Journal*, *120*(3), 480–506. https://doi.org/10.1086/707007

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511811678

Moore, T. J., Johnston, A. C., & Glancy, A. W. (2020). STEM integration: A synthesis of conceptual frameworks and definitions. In C. C. Johnson, M. Mohr-Schroeder, T. J. Moore, & L. English (Eds.), *Handbook of research on STEM education* (pp. 3–16). Routledge. https://doi.org/10.4324/9780429021381-2

National Center for Education Statistics. (2019). The Nation's Report Card: 2019 Mathematics Assessment. https://nces.ed.gov/nationsreportcard/mathematics/

National Center on Intensive Intervention. (n.d.). Academic Intervention. Retrieved from https://intensiveintervention.org

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

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. U.S. Department of Education.

National Research Council. (2009). *Mathematics Learning in Early Childhood: Paths Toward Excellence and Equity*. The National Academies Press. https://doi.org/10.17226/12519

Nelson, G., & McMaster, K. L. (2019). The effects of early numeracy interventions for students in preschool and early elementary: A meta-analysis. *Journal of Educational Psychology*, *111*(6), 1001–1022. https://doi.org/10.1037/edu0000334

Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities*, *51*(6), 523–539. https://doi.org/10.1177/0022219417714773

Next Generation Science Standards. (2013). *Next generation science standards: For states, by states*. The National Academies Press.

Parmar, R. S., Garrison, R., Clements, D. H., & Sarama, J. (2011). Measurement. In F. Fennell (Ed.), *Achieving fluency: Special education and mathematics* (pp. 197–218). National Council of Teachers of Mathematics.

Polikoff, M. S. (2012). The redundancy of mathematics instruction in U.S. elementary and middle schools. *The Elementary School Journal*, *113*(2), 230–251. https://doi.org/10.1086/667727

Powell, S. R. (2011). Solving word problems using schemas: A review of the literature. *Learning Disabilities Research & Practice*, *26*(2), 94–108. https://doi.org/10.1111/j.1540-5826.2011.00329.x

Powell, S. R., Cirino, P. T., & Malone, A. S. (2017). Child-level predictors of responsiveness to evidence-based mathematics intervention. *Exceptional Children*, *83*(4), 359–377. https://doi.org/10.1177/0014402917690728

Pressley, M., Graham, S., & Harris, K. (2006). The state of educational intervention research as viewed through the lens of literacy intervention. *British Journal of Educational Psychology*, *76*(Pt. 1), 1–19. https://doi.org/10.1348/000709905X66035

Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M. (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary. *Reading & Writing Quarterly*, *31*(3), 235–252. https://doi.org/10.1080/10573569.2015.1030995

Sarama, J., Clements, D. H., Baroody, A. J., Kutaka, T. S., Chernayavskiy, P., Shi, J., & Cong, M. (2021). Testing a theoretical assumption of a learning-trajectories approach in teaching length measurement to kindergartners. *AERA Open*, *1*, 1–15. https://doi.org/10.1177/23328584211026657

SAS Institute. (2016). *SAS/STAT 14.2 user's guide*. SAS Institute, Inc.

Satsangi, R., Bouck, E. C., Taber-Doughty, T., Bofferding, L., & Roberts, C. A. (2016). Comparing the effectiveness of virtual and concrete manipulatives to teach algebra to secondary students with learning disabilities. *Learning Disability Quarterly*, *39*(4), 24. https://doi.org/10.1177/0731948716649754

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

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

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 957–1009). National Council of Teachers of Mathematics.

Sutherland, M., Firestone, A. R., Doabler, C. T., & Clarke, B. (2020). Building conceptual understanding of linear measurement: Teaching students with mathematics learning disabilities. *Teaching Exceptional Children*, *53*(1), 70–78. https://doi.org/10.1177/0040059920919131

Swanson, E., Wanzek, J., Vaughn, S., Fall, A. M., Roberts, G., Hall, C., & Miller, V. L. (2017). Middle school reading comprehension and content learning intervention for below-average readers. *Reading & Writing Quarterly*, *33*(1), 37–53. https://doi.org/10.1080/10573569.2015.1072068

The Institute of Education Sciences, U.S. Department of Education and the National Science Foundation. (2013). *Common guidelines for education research and development*. The Institute of Education Sciences, U.S. Department of Education, National Science Foundation. https://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf

University of Texas System/Texas Education Agency. (2008). *Texas Early Mathematics Inventory–Progress Monitoring*.

Vasilyeva, M., Laski, E. V., Veraksa, A., & Bukhalenkova, D. (2020). Leveraging measurement instruction to develop kindergartners' numerical magnitude knowledge. *Journal of Educational Psychology*. Advance online publication. https://doi.org/10.1037/edu0000653

Vaughn, S., Capin, P., Scammacca, N., Roberts, G., Cirino, P., & Fletcher, J. M. (2019). The Critical Role of Word Reading as a Predictor of Response to Intervention. *Journal of Learning Disabilities*. Advance online publication. https://doi.org/10.1177/0022219419891412

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician*, *70*(2), 129–133. https://doi.org/10.1080/00031305.2016.1154108

What Works Clearinghouse. (2020). *Procedures and standards handbook* (Version 4.1). U.S. Department of Education, Institute of Education Sciences.

What Works Clearinghouse. (n.d.). What works in math? Retrieved from https://ies.ed.gov/ncee/wwc/FWW/Results?filters=,Math

Zucker, T., Carlo, M. S., Landry, S. H., Masood-Saleem, S. S., Williams, J. M., & Bhavsar, V. (2019). Iterative design and pilot testing of the developing talkers tiered academic language curriculum for pre-kindergarten and kindergarten. *Journal of Research on Educational Effectiveness*, *12*(2), 274–306. https://doi.org/10.1080/19345747.2018.1519623

Received July 3, 2020
Revision received July 31, 2021
Accepted August 13, 2021 ■