WestEd

Measuring What Matters

Considerations for Measuring the Implementation of High-Quality Instructional Materials in Mathematics

Rebecca Perry, Drew Nucci, Kirk Walters, David McKinney, Sarah Nielsen, and Maggie Hannan

February 2026

Suggested citation: Perry, R., Nucci, D., Walters, K., McKinney, D., Nielsen, S., & Hannan, M. (2026). *Measuring what matters: Considerations for measuring the implementation of high-quality instructional materials in mathematics*. WestEd.

WestEd is a nonpartisan, nonprofit organization that aims to improve the lives of children and adults at all ages of learning and development. We do this by addressing challenges in education and human development, increasing opportunity, and helping build communities where all can thrive. WestEd staff conduct and apply research, provide technical assistance, and support professional learning. We work with early learning educators, classroom teachers, local and state leaders, and policymakers at all levels of government.

For more information, visit WestEd.org. For regular updates on research, free resources, solutions, and job postings from WestEd, subscribe to the *E-Bulletin*, our semimonthly e-newsletter, at WestEd.org/subscribe.

Table of Contents

Executive Summary

High-quality instructional materials (HQIM)—a standards-aligned core curriculum and the publisher-developed resources designed to support teachers' use of the curriculum—are increasingly central to district efforts to improve mathematics instruction. Despite a national emphasis on HQIM adoption, the field still lacks common, public-facing information about how HQIM from different publishers are designed to support teaching and learning. Because curriculum development is often proprietary, a lack of access complicates the ability of researchers, educators, and policymakers to differentiate between programs in meaningful ways.

As such, much remains unknown about how HQIM are implemented in classrooms and experienced by students. For example, many mathematics materials might prioritize procedural fluency grounded in conceptual understanding, but individual materials might differ significantly from each other in terms of their embedded features and in what they emphasize when implemented with fidelity. These differences create measurement challenges for research on HQIM implementation and impact. In this report, a WestEd research team raises these measurement challenges, and we describe how we are attempting to address them in a nationwide study of HQIM implementation and impact being conducted from 2024 through 2027. This report is the first in a series of reports to be developed from this study.

Subsequent reports produced from this study will be grounded in the methodological approaches described herein. As we continue this research—further gathering, reviewing, summarizing, and sharing the collected data—we will also refine our methods to better capture implementation processes that are inherently dynamic and context dependent. We hope that providing the foundational information in this first report will allow us to streamline future reports for readers interested in the national HQIM movement to quickly digest important findings and new methodological considerations associated with HQIM implementation and impact.

Introduction

The choice of what curriculum to use is widely recognized as a critical lever for improving instruction and student achievement (D. Steiner, 2017; Koedel & Polikoff, 2017). In recent years, especially since the COVID-19 pandemic, there has been a growing emphasis on the use of what are known as high-quality instructional materials (HQIM) to improve student outcomes (e.g., Hartl & Riley, 2021; Livingston, 2023). In general, the term "HQIM" is used to refer to instructional materials that are closely aligned with educational standards and user-friendly for both teachers and students. Typically HQIM also feature evidence-based teaching methods and uphold a commitment to equity and inclusion, with supports to enable all students to access grade-level content (EdReports, n.d.-a; EdTrust-West, 2023). They may also include guidance to facilitate implementation, such as recommendations about initial training and ongoing professional development. The nationwide emphasis on HQIM is demonstrated by the fact that 15 states are collaborating with the Council of Chief State School Officers (CCSSO) Instructional Materials and Professional Development Network to support HQIM-aligned curriculum selection and professional learning (CCSSO, n.d., 2024), while others have launched state-level initiatives such as California's HQIM Learning Community (EdTrust-West, n.d.) and Louisiana's instructional materials rating system (Louisiana Department of Education, n.d.).

The EdReports rating system has played a big role in this shift for over 40 states since 2015 by providing independent assessments of instructional materials' quality and offering insight about promoting and selecting HQIM (EdReports, n.d.-b). Recent data show that 42 percent of sampled districts selected a high-quality, aligned core curriculum for middle school math (grades 6–8) and 47 percent of districts selected a high-quality, aligned core curriculum for K–5 math (Center for Education Market Dynamics, 2024). Use of HQIM such as Illustrative

Mathematics, Eureka Math, Carnegie Learning, and i-Ready Classroom Mathematics is increasing, while districts' use of poorly rated materials is decreasing.

However, because the concept of HQIM is still relatively new, there is still little evidence about the efficacy of HQIM in math instruction—little is known about how best to implement HQIM or their impact. Moreover, the overall increase in adoption of different HQIM may complicate impact studies and make it more difficult to detect treatment effects for one program in relation to others because of a potentially weaker service contrast between programs. However, there have been a few impact studies to date. A study of Desmos at the middle grades (McKinney et al., 2023) found a positive effect on middle grade math scores using just a basic measure of implementation usage. Another quasi-experimental study using HQIM adoption data found no statistically significant differences on mean student mathematics achievement in grade 6 for Illustrative Mathematics districts compared to similar districts using business-as-usual mathematics curricula (DePiper et al., 2023). A study of individual school districts (Cook et al., 2023) found that students in Illustrative Mathematics classrooms achieved higher gains on a computer-administered progress-monitoring assessment from the beginning to the end of the school year compared to their peers in a control group. More recently, our team found that Massachusetts schools adopting Illustrative Mathematics prior to the COVID-19 pandemic significantly outperformed comparable schools, particularly in grades 7 and 8, by the third year of implementation (Khanani et al., 2025). How these documented impacts relate to local implementation, context, and professional learning has not yet been rigorously studied.

Prior research suggests that contextual variables such as school leadership, teacher preparedness, student demographics, alignment of instructional systems, and school resources moderate curriculum efficacy (Borman et al., 2003/2016; Desimone et al., 2002), and curricula are likely to be more efficacious when combined with professional learning (Desimone, 2009; Hill et al., 2020; Hill & Papay, 2022). Context and professional learning matter insofar as they can shift teacher mindsets and practices. Teaching remains the "black box" of curriculum implementation because teachers add, modify, and omit important curricular design features based on their own attitudes, beliefs, and goals for student learning (Brown, 2009).

Implementation fidelity—the degree to which an educational intervention or program is delivered as intended by its developers—has drawn increasing attention since the founding of the Institute of Education Sciences (IES) in 2002, but some researchers have argued that implementation fidelity remains underexamined (Gresham, 2017). Its measurement ensures that the intervention's outcomes can be attributed to the program itself rather than to variations in delivery. High fidelity enables stronger attribution of outcomes to the intervention (Durlak & DuPre, 2008; O'Donnell, 2008), while lack of fidelity measurement undermines evaluation usefulness (Carroll et al., 2007; Peterson et al., 1982). For example, without measuring fidelity precisely, a program could be deemed ineffective not because of its inherent design but due to poor implementation. Understanding implementation fidelity also provides valuable insights for program development and refinement. Identifying elements that are consistently well implemented can highlight program strengths, while pinpointing areas with low fidelity can reveal where additional support or modification might be needed (Century et al., 2010).

Measuring implementation fidelity, especially for HQIM, is complex. As discussed in more detail later in this report, identifying which dimensions are core to the intervention and which need to be measured to assess fidelity is not always straightforward (Carroll et al., 2007; Century et al., 2010; D. L. Fixsen et al., 2009; Lynas & Hawkins, 2017). Similarly, a minimum implementation fidelity threshold that must be achieved is not always clearly specified (A. A. M. Fixsen et al., 2021; Bertram et al., 2015; D. L. Fixsen et al., 2009), and all educational interventions are implemented across a range of complex settings (Berman & McLaughlin, 1974; D. L. Fixsen & Blase, 2020; D. L. Fixsen et al., 2015; Durlak & DuPre, 2008) at differing stages of readiness (Dane & Schneider, 1998). Classroom-based interventions such as HQIM add further complexity to measuring fidelity because teachers must make professional decisions in complex, dynamic environments (Elmore & McLaughlin, 1988; Remillard & Heck, 2014). Professional learning can help reduce variability by building teachers' understanding of curriculum design and intended classroom interactions (Alicea et al., 2025). This support may include professional learning sessions provided by curriculum designers, coaching, professional learning communities (PLCs), or some combination thereof. Also, the curriculum itself can be "educative," with teacher guides that explain pedagogical intent and scaffold enactment (Davis & Krajcik, 2005). In an effort to support educators' professional learning, many districts engage in their own learning to internalize the curriculum's vision and use data collection tools to monitor and support implementation fidelity within the system.

Purpose and Research Questions

The primary purpose of the research methods described in this paper is to contextualize a series of efficacy studies comparing Illustrative Mathematics to other middle school curricula. These efficacy studies are part of a national study being conducted by WestEd from 2024 through 2027 that is focused on the implementation and impact of mathematics HQIM. Our research focuses particularly on Illustrative Mathematics—and other HQIM—in the middle grades because the Illustrative Mathematics 6–8 materials have been in place in U.S. schools longer than many other HQIM (and longer than Illustrative Mathematics at other grade levels) and therefore have had more opportunity to have been the focus of implementation over multiple years, hopefully in support of greater impact on student outcomes. Given that elementary and high school contexts are quite different for teachers, this report's research team did not anticipate being able to generalize findings beyond the middle grades, and we designed research to answer four primary questions, with a goal of exploring these questions to test and refine preliminary measures to capture critical implementation components at the middle school level. Our four main research questions are as follows:

- **Q1:** How do various interest holders—curriculum developers, district leaders, and teachers—communicate about quality implementation? In short, what is the intended curriculum, and what guidance do developers provide to educators involved in implementation?
- **Q2:** How do school districts understand curriculum adoption and implementation? What are school districts' challenges in implementation fidelity, and how do they support quality implementation?
- **Q3:** How do teachers characterize their HQIM implementation and the support they receive? How do students experience the implementation, and is their experience aligned with the teachers' characterization of implementation?
- **Q4:** What data displays and communication structures help instructional leaders support implementation and continuous improvement?

What follows is the first in a series of reports focused on answering these questions. This first report discusses the research design and measurement approaches for understanding HQIM implementation across a national sample of districts to be studied; subsequent reporting will share findings.

Framing

To answer the research questions about HQIM implementation, we drew on two central research frameworks—one about instructional materials and one about implementation fidelity.

How We Understand Instructional Materials

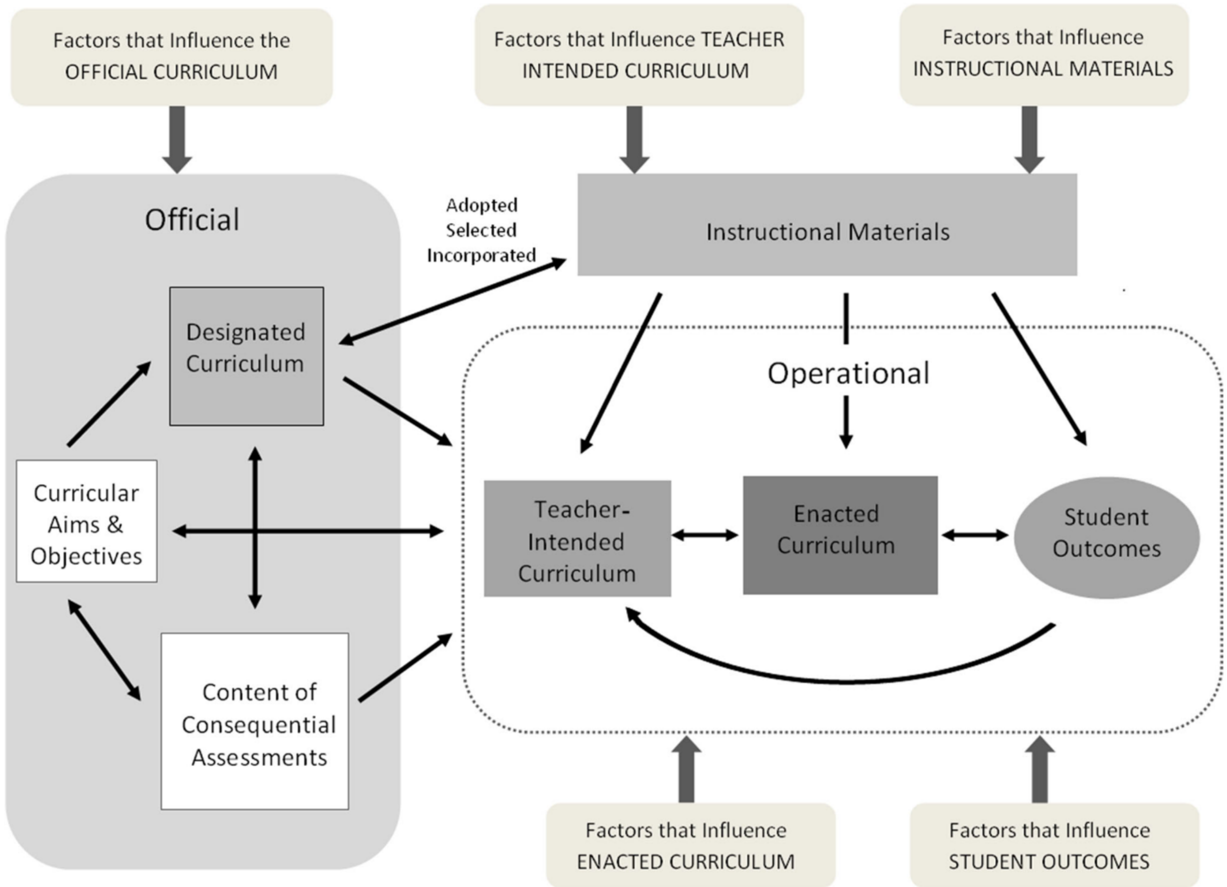
We conceptualize instructional materials (including digital resources) as resources that specify intended teacher and student interactions with mathematics (Rezat & Sträßer, 2012). However, these intended interactions seldom go as planned because teachers make a series of professional decisions to interpret and implement the materials in the complex environment of a classroom. Researchers (e.g., Remillard, 2018) have characterized this process as the translation from the “official curriculum” to the teachers’ “intended” curriculum to the curriculum as “enacted” with students.

Curriculum developers design the *official* curriculum as a tool to mediate activity, with representations of both concepts and procedures (Brown, 2009) and a vision of classroom interaction (Van Steenbrugge & Remillard, 2023). Teachers then add, modify, and omit curricular tasks and interactions based on their attitudes, beliefs, and goals for student learning (Brown, 2009). They analyze the materials, considering alignment with standards, benchmarks, and the diverse needs of their students, and they make decisions about how to best structure lesson plans, allocate time for various topics, and choose appropriate instructional strategies and resources. This is how they craft the *intended* curriculum, and this phase is heavily influenced by teachers’ knowledge, experiences, and beliefs about teaching, as well as by contextual factors such as district pacing guides, school policies, available resources, and class demographics.

However, students also have agency. Teachers must respond and adapt to students in the moment as lessons unfold and the curriculum becomes *enacted* (Remillard & Taton, 2013). During instruction, teachers must dynamically respond not only to pressures they feel to cover material and keep up with their district pacing guide but also to students' questions, misconceptions, and varying levels of understanding. These pressures often require on-the-spot modifications to the lesson tasks, activities, and interactions—for example, whether they cover all the questions provided by the curriculum or choose to go deeper in certain areas. Other adaptations could include altering the sequence of topics, varying instructional methods, or incorporating additional supports and materials. Teachers also assess the effectiveness of their instruction through formal and informal assessments, using the insights gained to further refine their teaching practices. Thus, the enacted curriculum is often a more personalized and flexible version of the intended curriculum, shaped by ongoing reflection, feedback, and the evolving classroom environment.

Figure 1 illustrates these translational processes (Remillard & Heck, 2014, p. 709), showing curriculum development and implementation as a dynamic system with two main sections: the official curriculum and the operational curriculum. The official curriculum encompasses three interconnected components—curricular aims and objectives, a designated curriculum, and content of consequential assessments—all influenced by external factors. The designated curriculum serves as a bridge to instructional materials, which are shaped by factors influencing both the official curriculum and teacher-intended curriculum. These materials undergo adaptation as they move into the operational phase. Within the operational curriculum, instructional materials flow to the teacher-intended curriculum, which then informs the enacted curriculum in the classroom, ultimately affecting student outcomes. The model emphasizes reciprocal relationships throughout: Student outcomes feed back to influence the enacted curriculum, the enacted curriculum influences the teacher-intended curriculum, and these operational elements connect back to the official curriculum components. External factors continuously shape all stages of the enactment process, as educators interpret the official curriculum in practice.

Figure 1. Transitions in the Use of Instructional Materials



Implementation Fidelity

This study considers implementation to be a reduction in the translational difference between the essential, intended design elements of the curriculum and its enactment in classrooms. Specifically, in this research study, we seek to operationalize the dosage, adherence, instructional quality, and student experience of the curriculum (Gage et al., 2020), also adding consideration for the local supports that must be activated to facilitate the delivery of any policy or program such as HQIM.

- **Dosage: how much a teacher uses the official curriculum**—A recent RAND survey found that half of all middle school math teachers use their own materials at least weekly (rather than using only the materials prescribed by the school or district), and 42 percent report that their principal encourages them to use whatever materials best meet their students' needs (Doan et al., 2023). These findings have powerful implications for teaching and learning with HQIM insofar as the lower dosage may limit the intended impact. In this study, we seek to understand teachers' coverage of lessons within a unit and the extent to which they are using the core curriculum versus other curricular resources.

- **Adherence: how closely teachers follow official curriculum resources**—A key element of implementation fidelity is the extent to which teachers adhere to the prescribed units, lessons, and sequence described by developers. Curriculum developers organize mathematical ideas into a sequence of units, lessons, and activities to guide student learning, often hierarchically. This sequencing is one way that today's HQIM differ from mathematics curricula of prior eras and from each other. In this study, we seek to understand teachers' pacing and ordering of units and the proportion of activities from the curriculum typically used in instruction. Each of these teacher decisions can impact student exposure to the HQIM content.
- **Instructional Quality: how closely instruction aligns with the intended interactions**—Because teachers "use new curricula in cursory ways" (D. M. Steiner & Kane, 2019), the quality of instruction is a vital piece of implementation fidelity. Quality is more difficult to define, given the complexity of classroom instruction generally, but in this study we aim to define guidelines about quality from how curriculum developers describe their intended classroom content interactions.
- **Student Experience: students' perceptions of mathematical opportunities to learn and interactions in class**—We see student experience of a curriculum as a vital element of implementation fidelity because of the last transition in *enacted* curriculum use. All developers prioritize curriculum coherence, and this is one of the metrics in the HQIM profiles of EdReports. However, classroom practice may not meet a coherence standard for implementation fidelity if teachers' and students' reports differ. As later sections discuss, this potential disconnect is why we created student-facing data instruments to triangulate teacher self-reports on their teaching practices.
- **Support**—We also consider teachers' implementation to be a contextual phenomenon. Instructional policy has to grapple with inherently incoherent demands (Russell & Bray, 2013), leaving it up to each district and school to "craft" (Honig & Hatch, 2004) the necessary amount of coherence to support leaders and teachers in their implementation efforts. Teachers' implementation decisions are informed by the professional learning and support opportunities afforded by their districts and schools, which in turn informs their own goals, attitudes, and beliefs as they choose to add, modify, or omit elements of the official curriculum. Clarity and alignment across professional learning and other instructional guidance levers, such as

materials and assessments, can be an especially powerful support for effective implementation (Coburn et al., 2016). Professional learning resources and supports can also vary dramatically across districts and remain largely hidden to external actors, making it challenging for districts who may be implementing the same HQIM to learn from each other about professional learning strategies and successes to support implementation.

Methodological Approach

We have been drawing on the instructional materials and implementation fidelity frameworks described as we design data collection methods for this study. We are collecting data from individuals and groups who are key to the curriculum enactment process: curriculum developers, district administrators, teachers, and students. An important goal of the study's multipronged data collection is to triangulate an understanding of what is happening locally in order to create a single implementation rating (e.g., low, mid, high) to include in the study's impact models. Below we describe our preliminary approaches to gathering data from each group. (Protocols for each of these methods are available upon request.)

Curriculum Review and Interviews With Curriculum Developers

Grounding our research in a thorough understanding of HQIM is important for clarifying what makes each set of materials "high quality" (Schwartz, 2025) and understanding the materials' similarities to and differences from each other in how they frame issues of adoption, implementation, and support. As we were in early discussions with districts for study recruitment, our team of math education experts reviewed the student and teacher materials for the math curricula adopted by those districts: Illustrative Mathematics, Carnegie Learning, i-Ready, and Savvas, all of which were rated highly on EdReports. Reviewing curriculum materials is important for defining the treatment and core program dimensions to operationalize fidelity (Carroll et al., 2007; Century et al., 2010; Moncher & Prinz, 1991). As these programs each attempt to serve similar purposes and draw on similar research-based goals and practices for mathematics education, program study and differentiation is a critical step for later being able to interpret results of the efficacy study (Gage et al., 2020).

Curriculum Review

Our research seeks to better understand the curricula by analyzing three categories of curricular features: the structure, the developer's intended classroom interactions, and curriculum-embedded educative features for teachers. To document the structural similarities and differences across curricula, we are creating a list of curriculum features (e.g., number of units and lessons per grade level, how mathematical concepts were clustered) and entering results from our review for each HQIM into a comparison spreadsheet. The team then collectively reviews and discusses the entries for each feature to ensure that we are calibrated in our understanding of how to characterize the HQIM features.

To understand the materials' educative features—those intended to support teachers' learning about instructional planning—we are studying the teacher guides and resources. All of the HQIM provide recommendations for teaching, with some variation in the justification provided for teachers to understand the basis for those recommendations. For example, one form of guidance embedded directly in all four HQIM reviewed so far is information about the five Practices for Orchestrating Productive Mathematical Discussions (M. S. Smith & Stein, 2011). Guidance from Illustrative Mathematics states the five practices and the curricular features that support them but does not expand on what each is and why they are important to implement. Carnegie Learning states the five practices and provides guiding questions for teachers to consider but does not say why the five practices support student learning nor how—for example, it does not say anything about teachers planning to sequence student ideas. The i-Ready materials state the practices and why they are important and provide guidance on how teachers should do the work, including examples. As with Illustrative Mathematics, the Savvas enVision teachers' guide states the five practices and the text features in the teachers' guide that could support teachers in anticipating or sequencing student ideas. Savvas explains why the practices are important and how to use the curriculum to implement each practice.

We are further reviewing how the HQIM show possible student solutions and what discussion questions they suggest to teachers to support student thinking and discourse. We are reviewing for teacher prompts and questioning because they have been shown to be such a central factor in students' classroom discourse and how much mathematics is learned (Boaler & Brodie, 2007), and the different sets of HQIM we have reviewed treat questioning quite differently.

Interviews With Curriculum Developers

We are interviewing HQIM developer teams about their materials and their expectations for implementation, drawing on prior research to inform the interview protocol development (e.g., Choppin et al., 2022; Hord et al., 2013; Huntley, 2009; McNaught et al., 2010). First, we ask each developer team to provide written responses to a series of 10 questions about their characterization of the core curriculum and its essential features, implementation fidelity thresholds (e.g., percentage of units, lessons, and activities), faithful adaptations, and system and teacher readiness suggestions. The protocol solicits developers' perspectives on the ways in which their curriculum scaffolded learning opportunities along the following four dimensions:

- opportunities to engage in **meaningful mathematical reasoning**, such as modeling real-world phenomena, solving novel and complex problems, solving problems with multiple solutions, and/or using a variety of strategies to solve problems
- opportunities to develop **procedural fluency** based on **conceptual understanding**
- opportunities to **demonstrate knowledge** and engage in mathematically **rigorous discourse** wherein students create **multiple representations** of their thinking, **justify their thinking**, critique/interrogate the thinking of others, and have opportunities to ask conceptual questions
- **tailored learning opportunities** wherein teachers adjust instruction and instructional activities based on formative assessment and the needs of the students, providing scaffolding for students who are English Learners and differentiation for students with varied learning needs

We are inquiring about the frequency with which essential teacher and student activities and interactions should be expected to occur in classrooms with adequate implementation. This information is meant to equip us to score each essential activity as one that would either promote or inhibit student learning opportunities. We also ask them to identify the minimum required lesson or unit components that should be completed to meet expectations for adequate implementation. A broader goal is to establish an implementation threshold for each program—a minimum characterization of what implementation should look like across districts and schools and classrooms that all use the same set of materials. Our assumption is that unless this threshold is clearly specified for district leaders, they might not be able in turn to communicate to teachers a clear expectation for the materials' usage.

After developers complete the spreadsheet, we review the responses and create a semistructured interview protocol for a 1-hour follow-up discussion with the development team to prompt elaboration on their responses. For the development teams that were not able to meet with us, our team completed protocol responses using available materials. In this case, the teacher and student materials more clearly specify information about essential features than about implementation threshold and allowable adaptations.

We added another layer of interviews with representatives of two certified provider organizations for the open source Illustrative Mathematics curriculum: Kiddom and Imagine Learning. We were aware that these intermediary organizations and the features offered through their learning management systems might serve as additional implementation guidance and therefore create additional complexity for district and school educators.

Data Collection With District Leaders

To understand how professional learning providers and district and site administrators operationalize and provide support for teachers' HQIM implementation, we organized three types of data collection, detailed in the following sections.

Interviews and Meetings With District Administrators

We have planned and begun carrying out two or three 60- to 90-minute structured meetings across each school year with a small implementation support team in each district. The initial meeting is designed to understand the district's history with and motivation for using the HQIM, supports for and challenges with implementation, and the way the district team thinks about and measures implementation fidelity (Cutbush et al., 2017). For example, because curriculum developers emphasize implementation "integrity" over strict fidelity, we want to understand how districts might operationalize this distinction. We also want to understand how districts set up systems to monitor HQIM implementation—for example, by using classroom observations, internal tools such as teacher surveys, or collaborative planning or analysis routines. We are also gathering district artifacts related to implementation, such as pacing guides and professional learning schedules.

Subsequent meetings are designed as continuous improvement discussions, being guided by district leaders' inquiry questions and using their own teacher and student data to touch on implementation successes and challenges. We are using these meetings to summarize and share district-specific aggregate teacher and student survey data and to engage leaders in discussions to identify actionable next steps for their implementation. We also are seeking to understand how closely our data match their other sources of data about curriculum implementation and whether there are additional implications for professional learning supports.

Professional Learning Questionnaire

To focus on curriculum-based professional learning (CBPL), we created a questionnaire that solicits information about professional learning in each district and its schools. Following Alicea and colleagues (2025, p. 4), we defined CBPL as

> **professional learning that supports the implementation and sustained use of specific curriculum to ensure teachers' effective use of these materials in their classrooms. It aligns with the curriculum's content and instructional strategies, enabling teachers to provide instruction to students and facilitate learning with integrity to what was designed while adapting to the needs of their students.**

The questionnaire includes four sections: (a) basic facts about the district, including the number of middle school math teachers and how many hours of professional learning per teacher is built into the union agreement; (b) types of professional learning structures offered in the district, including PLCs and summer institutes, and the average monthly time that educators in the district received professional learning through each structure; (c) determination of whether the district's professional learning is aligned with 11 CBPL features, including whether it is individualized or student focused; and (d) ratings for CBPL and HQIM curriculum implementation for each school in the district. We designed the questionnaire to enable us to characterize the professional learning supports and identify best practices in districts where implementation is going well or where student achievement seems to be rising over time.

Site Administrator Focus Groups

Because prior implementation research has highlighted principal leadership as essential to supporting curriculum use, we also want to hear directly from site administrators. Beginning in the 2024–25 school year as districts were recruited into the study and through the 2026–27 end of the study, we are holding annual 1-hour focus groups with district principals to get their input on the mathematics curriculum and perceived alignment of the curriculum with the district's vision, their perspective on the teachers' use of the curriculum to develop the teachers' instructional practices, and information on the professional learning provided to support educators in learning to implement the curriculum. We are asking, for example, how districts support site leaders in building their own capacity for instructional leadership to support curriculum implementation—through leadership training, calibration meetings, or strategic walk-through participation.

Teacher Surveys

While direct observation is often considered preferable for understanding classroom instruction and measuring fidelity (Gresham, 2017), we have been keenly aware of the heavy burden on practitioners of focusing attention on pandemic recovery and learning loss in mathematics, particularly as this study was getting underway. We therefore chose methods that would be minimally burdensome. In addition, our aim of understanding HQIM implementation across a range of district and state contexts in a national study made the prospect of direct observation logistically and fiscally challenging (White & Klette, 2023). Because districts use different programs and different local measures of implementation fidelity, we cannot rely on locally gathered fidelity data, nor can we place the additional burden on district staff of learning how to use a common fidelity measure. Acknowledging the challenges related to the reliability of self-report data (Lira et al., 2022) and the need for frequent self-reports of fidelity (Gage et al., 2020), we have opted for teacher self-reporting through teacher surveys as the least burdensome and most reliable strategy for learning about how teachers were using the HQIM. We have designed teacher surveys to be administered three times annually during the 2024–2027 study period. Data from teacher surveys, aggregated to create individual implementation measures, could also be aggregated to the school level and compared with the school-level ratings provided by district leaders via the professional learning questionnaire.

Where possible, we are including in the teacher survey items from nationally validated surveys (e.g., Craven, 2019; Schweig et al., 2023) to be able to assess differences in responses between our intended national study sample and a nationally representative sample on topics such as the frequency of curriculum usage, the use of supplemental resources (including those that are self-made) alongside the core curriculum, and the extent of and reasons for curriculum modification or supplementation. We have added additional survey questions to the first and third survey administrations, asking about teachers' professional learning supports, beliefs about math teaching, and individual demographic characteristics. The survey is designed to take teachers no longer than 30 minutes per administration and be completed by all participating grades 6–8 teachers. Table 1 shows the primary categories of the teacher survey questions and examples of how we operationalize fidelity and professional learning support through the teacher survey items. Subsequent sections describe our initial analysis plans for developing constructs derived from the survey responses.

Table 1. Primary Dimensions of Implementation Fidelity and How the Teacher Survey Operationalizes Them

Dimension of implementation fidelity	As operationalized in teacher survey
Dosage is an index of the number of sessions implemented, the length of each session, and/or the frequency of components implemented. **(Use of this dimension is contingent on there being a dosage recommendation provided by developers.)**	Units teachers taught in whole or in part since the beginning of the 2024/25 school year The approximate percentage of mathematics class time in which teachers use their core curriculum with a focal class
Adherence is the extent to which program components are delivered as prescribed.	*[For most recently completed unit]—extent of teachers' use of the core curriculum, extent to which teachers pick from the core curriculum and skip parts, extent to which teachers supplement the core curriculum, extent to which teachers modify the core curriculum Additional curriculum materials teachers use regularly Extent to which teachers skip lessons within a unit
Instructional Quality focuses on the qualitative aspects of program delivery.	*How often teachers engage in each of the [listed] classroom instructional practices *How often teachers' students do each of the [listed] classroom instructional practices in class
Support focuses on the teachers' understanding of the school and district support for their implementation.	Number of hours teachers spent on various professional learning activities (e.g., schoolwide or districtwide PLCs focused on mathematics) Teachers' ratings of usefulness of each [listed] professional learning activity Ways in which teachers typically plan units or lessons for a focal class—by themselves, with another teacher or teachers, with an instructional coach or other school-based instructional leader

***Note*. Each item denoted with an asterisk (*) could fit in multiple fidelity categories.**

Curriculum Dosage

We are interested in documenting how much of their core curriculum teachers use each year and whether teachers move through the curriculum at the pace recommended by curriculum developers and their district. To assess these points, we are using district pacing guides to define what is "on track" with suggested pacing for a given time point during the year and to compare teachers' reports of their unit coverage to date. This allows us to ascertain at each teacher survey time point (fall, winter, spring) whether teachers are on track, ahead of pacing, or behind.

Adherence

Several of the survey questions provide insight into teachers' adherence to the curriculum. For example, our survey replicates the following question from the 2018 National Survey of Science and Mathematics Education (Craven, 2019).

> Please indicate the extent to which you did each of the following while teaching your last completed unit. (Select one response on each row.) [Options range from "Not at all" (1) to "To a great extent" (5).]
>
> a. I used the core curriculum to guide the structure and content emphasis of the unit.
>
> b. I picked what is important from the core curriculum and skipped the rest.
>
> c. I incorporated activities from other sources to supplement what the core curriculum was lacking.
>
> d. I modified activities from the core curriculum.

We will use responses from this question to create a single adherence score for each teacher, examining patterns in the data to create cut scores for low, moderate, and high adherence. Individual scores for teacher adherence can be aggregated to the school or district level to understand cross-site variation. Teachers' responses about supplementing and skipping may be difficult to interpret without qualitative understanding of common district practices. For example, a teacher might report that they were supplementing their use of Illustrative Mathematics if they were using Amplify Desmos, a separate program that builds on and adds to the core curriculum of Illustrative Mathematics. In this case, their adherence score would be lower even though they were essentially using similar instructional materials. The ability to use our knowledge of teachers' instructional contexts obtained in meetings with district staff is one of the key values of our mixed-methods research approach, enabling us to more clearly specify the quantitative measures that will be used in the impact analyses.

We hope to also examine the relationship between the measures of dosage and adherence to understand the extent to which our fidelity measures provided a consistent implementation picture for individual teachers. For example, clustering of teachers at the lower ends of both the dosage and adherence measures may signal lower curriculum fidelity.

Instructional Quality

Unlike other studies that use individual fidelity instruments, each of which is aligned with a different curriculum (e.g., Agodini et al., 2010), our study is developing a single instrument for use *across* programs to enable comparison and contrast. For the quality dimension, we began survey development by identifying the essential pedagogical features in the curriculum as determined by our materials review and discussions with developers. These pedagogical intentions fall into a few key categories: coherence, cognitive demand, collaboration and discourse, formative assessment, and differentiation. We used existing valid and reliable survey instruments (e.g., Carney et al., 2015; Clark et al., 2014; Doan et al., 2023; Horizon Research, 2018; T. Smith et al., 2019) aligned with these categories and reviewed items for clarity, alignment with the HQIM, and redundancy and adapted items slightly only when necessary. Importantly, we want to ask teachers to report on both their own actions and their students' actions. So, for example, we ask teachers how often they "facilitate discussion of the connections between various models and/or strategies," but we also ask them how often their students "discuss each other's strategies for solving a problem."

We aim to try two initial approaches for analyzing these survey items. First, we may be able to categorize survey items into multiple groups: those emphasized more in one set of HQIM or another (e.g., Illustrative Mathematics or Carnegie Learning), those emphasized in more than one set of HQIM, or those that are common features of middle school mathematics classrooms but not explicitly emphasized in any HQIM. With these categorizations, we might expect to see teachers' more frequent use of features aligned with Illustrative Mathematics in the Illustrative Mathematics districts and more frequent use of features from other HQIM in districts adopting those other HQIM. We would expect to see fewer differences across districts for the curriculum features common to multiple sets of HQIM or emphasized by none. The second approach will involve conducting a factor analysis of the instructional practices to identify latent variables and examining patterns across districts or across different HQIM.

Professional Learning Support

Given that quantifying professional learning for individual teachers would be challenging because learning supports could be provided or sought at the district, school, and/or individual levels, we are using a professional learning questionnaire, as described above, to gather information on the district and school supports, and we are asking questions of individual teachers to gather additional information for the teacher layer.

To determine the amount of professional learning teachers have experienced, we ask teachers to report on how many hours of different types of professional learning they have received since the beginning of the school year. The survey adapted an item from a prior national survey to ask teachers about hours of formal professional learning offered by the school or district, mathematics coaching, nonevaluative walk-throughs or classroom observations of instruction, schoolwide or districtwide PLCs focused on mathematics, meetings or conferences offered by professional associations or networks (e.g., the National Council of Teachers of Mathematics), and online teacher communities aside from those at the school or district. We hypothesize that districts that have been implementing Illustrative Mathematics over longer periods might have teachers who report more total hours of professional learning because the district has found benefit in significant CBPL. We also ask teachers to report on the usefulness of each of these types of professional learning.

Another factor considered in operationalizing the professional learning measure is the value of professional collaboration in the learning process (Vescio et al., 2008). Two of the presumed benefits of HQIM are that the materials will save time for individual teachers by providing a starting place for planning high-quality mathematics units and lessons and will reduce variation and increase implementation fidelity among teachers within a school or district setting. Given the documented power of teacher collaboration, we expect coaching and PLCs to be two of the most frequent forms of professional learning reported, and we hypothesize that districts further into their curriculum implementation might have created more collaborative learning structures to better support their teachers (Hill & Papay, 2022). Data on these factors might also provide confirmatory information on which districts are emphasizing which professional learning strategies, enabling us to draw stronger conclusions about successful implementation supports across our sample. We also have considered the value of collaboration for teachers during lesson planning. We are asking teachers to report on several aspects of their planning process, including the amount of time they devote to planning lessons and units and the extent to which their lesson planning is a collaborative activity. We are considering, for example, whether teachers report planning mostly by themselves even in districts where PLC time has been provided.

To understand how well teachers feel their school or district professional learning has prepared them, we also ask questions about the activities and quality of the professional learning supports provided (e.g., the extent to which the professional learning is curriculum aligned and collaborative or involves analyzing student data) and whether teachers believe the strategies learned during professional learning lead to improved content knowledge, instruction, and understanding of how students learn math.

Student Surveys

As discussed above, our view of what goes into curriculum enactment and implementation fidelity includes students. To understand students' perspectives on their math classroom experiences, we developed a short survey (it takes no longer than 15 minutes to complete) for administering to students at the beginning and end of each school year. Survey development began much like the development of the teacher surveys: We used the same categories of coherence, cognitive demand, collaboration and discourse, formative assessment, and differentiation, and we used data from the curriculum review and developer discussions to characterize the students' intended interactions with mathematics, the teacher, and each other. We then identified items for these constructs from existing valid and reliable survey instruments (e.g., Barbieri & Miller-Cotto, 2021; Sinclair et al., 2023; T. Smith et al., 2019).

Mathematical Experiences

One set of questions on the student survey focuses on students' experiences of their mathematics classroom, asking questions aligned with several that we asked teachers about coherence, collaboration and discourse, and mathematical rigor. As we plan to do with the teacher survey data, we will conduct an exploratory factor analysis of student responses to these questions to identify latent variables such as a discourse scale. We can then review these scales by HQIM and by district to understand patterns of variation. For items that are asked of both teachers and students or latent constructs that are common for both groups, we will use simple correlation plots to look for agreement between teachers' and students' reports of classroom experiences.

Sense of Belonging and Self-Efficacy

We also included items on the student survey to ask about students' sense of belonging in the classroom because this is a significant predictor of motivation (as described in Barbieri & Miller-Cotto, 2021). Nine additional survey items were also included to replicate a reliable self-efficacy scale (Sinclair et al., 2023, 2025). Student self-perceptions of mathematics and themselves as math learners have important implications for their math learning and achievement, especially for students of color (e.g., Brisson et al., 2017; Martin et al., 2017; Rittle-Johnson et al., 2021), and HQIM curriculum developers are now attending to the goals of improving these student outcomes with specific curriculum features. For example, Illustrative Mathematics provides a *Building a Mathematical Classroom Community Plan* to bolster student feelings of safety, success, and ownership. This plan includes guidelines on student prompts and ways to record student thinking so that it is visible during instruction and can be referred to during the lesson.

Teacher Focus Groups

One final data collection strand involves engaging teachers in each district in annual 1-hour focus groups so that we gain information that is more detailed about implementation. We have designed focus groups to address three question categories: experiences with and perceptions of the mathematics curriculum, implementation guidance and support from district leadership about following the curriculum, and common modifications or changes to the curriculum. Not knowing how interested teachers might be in participating in these focus groups, we designed questions to triangulate information from other sources in the event of low teacher interest.

Conclusions and Next Steps

HQIM implementation data and analyses are rarely publicly available and yet are critical to understanding the impact of HQIM on student outcomes. We hope that our study and its methods described in this initial paper will enable us to begin to peel the onion of complex implementation processes. As we continue to recruit sufficient districts for a rigorous national study, we anticipate that there will be challenges in the data collection process at each of the four curriculum enactment stages. For example, one limitation we have already encountered is an inability to get direct input on essential curriculum features and implementation fidelity from all four of the curriculum development teams we contacted. With additional input, our teacher survey could have better reflected the specifics of those HQIM and enabled us to think more broadly about the survey questions we are asking teachers to report on.

Recruiting districts using multiple different sets of HQIM is another limitation we may need to overcome in our data collection and analysis. Recruiting a balanced sample of participants from Illustrative Mathematics and non–Illustrative Mathematics schools and gathering self-report implementation data from a large group of teachers in each district will both be quite important for a strong impact study. Additionally, finding time with busy district leaders for them to share how they think about and communicate to teachers about implementation fidelity is another expected challenge in this research.

Even with these potential limitations and challenges, we expect that we will learn a lot about how to conduct this kind of research and the conditions within U.S. education that help or hinder research and shared practice related to implementing HQIM. As our study continues, we look forward to reporting on what we are learning to inform both our district partners and the field of math education more broadly.

References

Agodini, R., Harris, B., Thomas, M., Murphy, R., & Gallagher, L. (2010). *Achievement effects of four early elementary school math curricula: Findings for first and second graders* (NCEE 2011-4001). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/2025/01/20114001-pdf

Alicea, S., Davis, C., Foster, E., Hornak, R., Hyler, M., Morrison, A., Papay, J., Richardson, J., Schwartz, N., & Wechsler, M. (2025). *Defining curriculum-based professional learning: Building a common language*. Research Partnership for Professional Learning. https://rpplpartnership.org/internal-resource/plpositionpaper/

Barbieri, C. A., & Miller-Cotto, D. (2021). The importance of adolescents' sense of belonging to mathematics for algebra learning. *Learning and Individual Differences*, *87*, Article 101993. https://doi.org/10.1016/j.lindif.2021.101993

Berman, P., & McLaughlin, M. W. (1974). *Federal programs supporting educational change, vol. 1: A model of educational change*. Rand Corp. https://files.eric.ed.gov/fulltext/ED099957.pdf

Bertram, R. M., Blase, K. A., & Fixsen, D. L. (2015). Improving programs and outcomes: Implementation frameworks and organizational change. *Research on Social Work Practice*, *25*(4), 477–487. https://doi.org/10.1177/1049731514537687

Boaler, J., & Brodie, K. (2007). *The importance, nature, and impact of teacher questions*. Stanford University.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2016). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, *73*(2), 125–230. https://doi.org/10.3102/00346543073002125 (Original work published 2003)

Brisson, B. M., Dicke, A.-L., Gaspard, H., Häfner, I., Flunger, B., Nagengast, B., & Trautwein, U. (2017). Short intervention, sustained effects: Promoting students' math competence beliefs, effort, and achievement. *American Educational Research Journal*, *54*(6), 1048–1078. https://doi.org/10.3102/0002831217716084 (Original work published 2017)

Brown, M. W. (2009). The teacher–tool relationship: Theorizing the design and use of curriculum materials. In J. T. Remillard, B. A. Herbel-Eisenmann, & G. M. Lloyd (Eds.), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 17–36). Routledge. https://doi.org/10.4324/9780203884645

Carney, M. B., Brendefur, J. L., Hughes, G. R., & Thiede, K. (2015). Developing a mathematics instructional practice survey: Considerations and evidence. *Mathematics Teacher Educator*, *4*(1), 93–118. https://doi.org/10.5951/mathteaceduc.4.1.0093

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*(1), Article 40. https://doi.org/10.1186/1748-5908-2-40

Center for Education Market Dynamics. (2024). *K–8 math curriculum quality: The state of district-led selection*. https://www.cemd.org/reportviewer/?pdfUrl=https://www.cemd.org/wp-content/uploads/2024/10/CEMD_K-8-Math-Curriculum-Quality_YOY_Fall2024.pdf&pdfName=K-8%20Math%20Quality%202024

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199–218. https://doi.org/10.1177/1098214010366173

Choppin, J., Roth McDuffie, A., Drake, C., & Davis, J. (2022). The role of instructional materials in the relationship between the official curriculum and the enacted curriculum. *Mathematical Thinking and Learning*, *24*(2), 123–148. https://doi.org/10.1080/10986065.2020.1855376

Clark, L. M., Neumayer DePiper, J., Frank, T. J., Nishio, M., Campbell, P. F., Smith, T. M., Griffin, M. J., Rust, A. H., Conant, D. L., & Choi, Y. (2014). Teacher characteristics associated with mathematics teachers' beliefs and awareness of their students' mathematical dispositions. *Journal for Research in Mathematics Education, 45*(2), 246–284. http://doi.org/10.5951/jresematheduc.45.2.0246

Coburn, C. E., Hill, H. C., & Spillane, J. P. (2016). Alignment and accountability in policy design and implementation: The Common Core State Standards and implementation research. *Educational Researcher*, *45*(4), 243–251.

Cook, M. A., Eisinger, J., & Ross, S. M. (2023). *Impact evaluation of Imagine Learning Illustrative Mathematics in Fort Zumwalt School District*. Johns Hopkins School of Education: Center for Research and Reform in Education.

Council of Chief State School Officers. (n.d.). *CCSSO's High-Quality Instructional Materials and Professional Development (IMPD) Network*. https://learning.ccsso.org/high-quality-instructional-materials

Council of Chief State School Officers. (2024). *Impact of the CCSSO IMPD Network*. https://learning.ccsso.org/impact-of-the-ccsso-impd-network-2024

Craven, L. M. (2019). *2018 NSSME+: Status of middle school mathematics*. Horizon Research. https://www.horizon-research.com/NSSME/wp-content/uploads/2019/05/2018-NSSME-Status-of-Middle-School-Math.pdf

Cutbush, S., Gibbs, D., Krieger, K., Clinton-Sherrod, M., & Miller, S. (2017). Implementers' perspectives on fidelity of implementation: "Teach every single part" or "Be right with the curriculum"? *Health Promotion Practice*, *18*(2), 275–282. https://doi.org/10.1177/1524839916672815

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, *18*(1), 23–45. https://doi.org/10.1016/S0272-7358(97)00043-3

Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, *34*(3), 3–14. https://doi.org/10.3102/0013189X034003003 (Original work published 2005)

DePiper, J. N., McKinney, D., Perry, R., Bartel, A., Huang, M., Kim, D., Kumar, K., & Thomas, R. (2023). *An exploratory study to understand the impact of the Illustrative Mathematics 6–8 curriculum*. WestEd.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140 (Original work published 2009)

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, *24*(2), 81–112. https://doi.org/10.3102/01623737024002081 (Original work published 2002)

Doan, S., Steiner, E. D., & Woo, A. (2023). *State of the American Teacher Survey: 2023 technical documentation and survey results*. Rand. https://www.rand.org/pubs/research_reports/RRA1108-7.html

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

EdReports. (n.d.-a). *Frequently asked questions*. https://edreports.org/about/faqs

EdReports. (n.d.-b). *Celebrating a decade of EdReports impact*. https://edreports.org/impact/edreports-10th-anniversary

EdTrust-West. (n.d.). *California High-Quality Instructional Materials Learning Community: Resource page*. https://west.edtrust.org/california-high-quality-instructional-materials-learning-community-resource-page/

EdTrust-West. (2023). *Defining high-quality instructional materials for mathematics: Centering the assets and needs of Multilingual Learner and English Learner students*. https://west.edtrust.org/wp-content/uploads/2024/01/Defining-High-Quality-Instructional-Materials-for-Mathematics.pdf

Elmore, R. F., & McLaughlin, M. W. (1988). *Steady work: Policy, practice, and the reform of American education* (ED 296 020). National Institute of Education.

Fixsen, A. A. M., Aijaz, M., Fixsen, D. L., Burks, E., & Schultes, M.-T. (2021). *Implementation frameworks: An analysis*. Active Implementation Research Network. https://static1.squarespace.com/static/6516b49bc5e0e32ff9cd1e2f/t/65b8dd3baa8f0b421c373be4/1706614076244/AIRN-AFixsen-FrameworksAnalysis-2021.pdf

Fixsen, D. L., Blase, K., Metz, A., & Van Dyke, M. (2015). Implementation science. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 695–702). Elsevier.

Fixsen, D. L., & Blase, K. A. (2020). Active implementation frameworks. In P. Nilsen & S. A. Birken (Eds.), *Handbook on implementation science* (pp. 62–87). Edward Elgar Publishing. https://www.e-elgar.com/shop/usd/handbook-on-implementation-science-9781788975988.html?srsltid=AfmBOooCCaVYcKSha4J0J4ALaFn9mHM58M8VgI586GHNAxKcbOXkDvUu

Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice*, *19*(5), 531–540. https://doi.org/10.1177/1049731509335549 (Original work published 2009)

Gage, N., MacSuga-Gage, A., & Detrich, R. (2020). *Fidelity of implementation in educational research and practice*. The Wing Institute. https://www.winginstitute.org/systems-program-fidelity

Gresham, F. M. (2017). Features of fidelity in schools and classrooms: Constructs and measurement. In G. Roberts, S. Vaughn, S. N. Beretvas, & V. Wong (Eds.), *Treatment fidelity in students of educational intervention* (pp. 22–38). Routledge.

Hartl, S., & Riley, C. (2021). *High-quality curriculum is a transformation tool for equity*. ASCD. https://www.ascd.org/el/articles/high-quality-curriculum-is-a-transformation-tool-for-equity

Hill, H. C., Lynch, K., Gonzalez, K. E., & Pollard, C. (2020). Professional development that improves STEM outcomes. *Phi Delta Kappan*, *101*(5), 50–56. https://doi.org/10.1177/0031721720903829

Hill, H. C., & Papay, J. P. (2022). *Building better PL: How to strengthen teacher learning*. Research Partnership for Professional Learning. https://annenberg.brown.edu/sites/default/files/rppl-building-better-pl.pdf

Honig, M. I., & Hatch, T. C. (2004). Crafting coherence: How schools strategically manage multiple, external demands. *Educational Researcher*, *33*(8), 16–30.

Hord, S. M., Stiegelbauer, S. M., Hall, G. E., & George, A. A. (2013). *Measuring implementation in schools: Innovation configurations*. Southwest Educational Development Laboratory. https://www.air.org/sites/default/files/CBAM-Innovation-Configurations.pdf (Original work published 2006)

Horizon Research. (2018). *National Survey of Science and Mathematics Education (NSSME+) Mathematics Teacher Questionnaire*. https://www.horizon-research.com/NSSME/wp-content/uploads/2018/12/2018-NSSME-MTQ-Formatted.pdf

Huntley, M. A. (2009). Measuring curriculum implementation. *Journal for Research in Mathematics Education*, *40*(4), 355–362.

Khanani, N., Zabala, D., Stackhouse, S., Gu, J., & Walters, K. (2025). *Impact of the Illustrative Math Curriculum on math achievement*. WestEd. https://www.wested.org/resource/impact-of-the-illustrative-math-curriculum-on-math-achievement/

Koedel, C., & Polikoff, M. (2017). Learning curve analysis and measurement of teacher effectiveness. *Economics of Education Review*, *61*, 280–291.

Lira, B., O'Brien, J. M., Peña, P. A., Galla, B. M., D'Mello, S., Yeager, D. S., Defnet, A., Kautz, T., Munkacsy, K., & Duckworth, A. L. (2022). Large studies reveal how reference bias limits policy applications of self-report measures. *Scientific Reports*, *12*, Article 19189. https://doi.org/10.1038/s41598-022-23373-9

Livingston, J. (2023, September 20). Getting high-quality math curriculum to the students who need it most. *The 74*. https://www.the74million.org/article/getting-high-quality-math-curriculum-to-the-students-who-need-it-most/

Louisiana Department of Education. (n.d.). *Instructional materials reviews*. https://doe.louisiana.gov/school-system-leaders/instructional-materials-reviews

Lynas, J., & Hawkins, R. (2017). Fidelity in school-based child sexual abuse prevention programs: A systematic review. *Child Abuse & Neglect*, *72*, 10–21. https://doi.org/10.1016/j.chiabu.2017.07.003

Martin, D. B., Anderson C. R., & Shah, N. (2017). Race and mathematics education. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 607–636). National Council of Teachers of Mathematics.

McKinney, D., Strother, S., Walters, K., & Schneider, S. (2023). *The effect of Desmos math curriculum on middle school mathematics achievement in nine states*. WestEd.

McNaught, M. D., Tarr, J. E., & Sears, R. (2010, April 30–May 4). *Conceptualizing and measuring fidelity of implementation of secondary mathematics textbooks: Results of a three-year study* [Paper presentation]. Annual Meeting of the American Educational Research Association, Denver, CO, United States. https://files.eric.ed.gov/fulltext/ED510115.pdf

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. *Clinical Psychology Review*, *11*(3), 247–266.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research, 78*(1), 33–84. https://doi.org/10.3102/0034654307313793

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, *15*(4), 477–492.

Remillard, J. (2018). Mapping the relationship between written and enacted curriculum: Examining teachers' decision making. In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu (Eds.), *Invited lectures from the 13th International Congress on Mathematical Education* (ICME-13 Monographs). Springer. https://doi.org/10.1007/978-3-319-72170-5_27

Remillard, J. T., & Heck, D. J. (2014). Conceptualizing the curriculum enactment process in mathematics education. *ZDM Mathematics Education*, *46*, 705–718. https://doi.org/10.1007/s11858-014-0600-4

Remillard, J. T., & Taton, J. (2013, April 17–20). *Design arcs and in-the-moment design decisions* [Paper presentation]. The Research Presession of the Annual Meeting of the National Council of Teachers of Mathematics, Denver, CO, United States.

Rezat, S., & Sträßer, R. (2012). From the didactical triangle to the socio-didactical tetrahedron: Artifacts as fundamental constituents of the didactical situation. *ZDM Mathematics Education*, *44*, 641–651. https://doi.org/10.1007/s11858-012-0448-4

Rittle-Johnson, B., Farran, D. C., & Durkin, K. L. (2021). Marginalized students' perspectives on instructional strategies in middle-school mathematics classrooms. *The Journal of Experimental Education*, *89*(4), 569–586. https://doi.org/10.1080/00220973.2020.1728513

Russell, J. L., & Bray, L. E. (2013). Crafting coherence from complex policy messages: Educators' perceptions of special education and standards-based accountability policies. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, *21*, 1–22.

Schwartz, S. (2025, April 2). What makes curriculum "high-quality"? *Education Week*. https://www.edweek.org/teaching-learning/what-makes-curriculum-high-quality/2025/04

Schweig, J., Pandey, R., Grant, D., Kaufman, J. H., Steiner, E. D., & Seaman, D. (2023). *American mathematics educator survey: 2023 technical documentation and survey results*. Rand.

Sinclair, H. C., Kim, Y. K., & Terrio, S. R. (2025). *The math mind measures: A user guide*. Social Research and Evaluation Center at the Louisiana State University College of Human Sciences and Education.

Sinclair, H. C., Rhodes, J., Robison, S., & Armour, N. (2023, May). *The math mind: Measuring motivation to engage with and persist in math* [Paper presentation]. Gates Foundation Advancing Innovative Math Solutions (AIMS) Collaboratory Convening, Seattle, WA, United States.

Smith, M. S., & Stein, M. K. (2011). *5 practices for orchestrating productive mathematics discussions*. National Council of Teachers of Mathematics; Corwin.

Smith, T., Walters, K., Griffin, M., Lennon, V., & Sanders, Z. (2019). *The better math teaching network: Lessons learned from Year 3*. Nellie Mae Education Foundation. https://nmefoundation.org/better-math-teaching-network-lessons-learned-from-year-3/

Steiner, D. (2017). *Curriculum research: What we know and where we need to go*. StandardsWork. https://standardswork.org/wp-content/uploads/2024/06/sw-curriculum-research-report-fnl.pdf

Steiner, D. M., & Kane, T. J. (2019). Don't give up on curriculum reform just yet. *Education Week*, *38*(28), 26–27.

Van Steenbrugge, H., & Remillard, J. T. (2023). The multimodality of lesson guides and the communication of social relations. *ZDM Mathematics Education*, *55*, 579–595. https://link.springer.com/article/10.1007/s11858-023-01479-2

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, *24*(1), 80–91.

White, M., & Klette, K. (2023). What's in a score? Problematizing interpretations of observation scores. *Studies in Educational Evaluation*, *77*, Article 101238. https://doi.org/10.1016/j.stueduc.2023.101238