



Structured Reporting Guidelines for Classroom Intervention Research

Heather C. Hill
Harvard University

Kathleen Lynch
University of Connecticut

Joshua R. Polanin
American Institutes for Research

Inconsistent reporting of critical facets of classroom interventions and their related impact evaluations hinders the field's ability to describe and synthesize the existing evidence base. In this essay, we present a set of reporting guidelines intended to steer authors of classroom intervention studies toward providing more systematic reporting of key intervention features and setting-level factors that may affect interventions' success. The guidelines were iteratively developed using recommendations and feedback from scholars active in conducting and synthesizing classroom intervention research. This effort aims to open wider the 'black box' in classroom research, communicating key information with more precision and detail to practitioners and future researchers, and permitting the field to more efficiently accumulate and synthesize findings on classroom interventions, determining what works, for whom, and under what conditions.

VERSION: August 2024

Suggested citation: Hill, Heather C., Kathleen Lynch, and Joshua R. Polanin. (2024). Structured Reporting Guidelines for Classroom Intervention Research. (EdWorkingPaper: 24-1022). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/50m5-8x28>

Structured Reporting Guidelines for Classroom Intervention Research

Heather C. Hill^a, Kathleen Lynch^b, and Joshua R. Polanin^c

^a Harvard Graduate School of Education

^b Neag School, University of Connecticut

^c American Institutes for Research

Author Note

Correspondence concerning this article should be addressed to Heather Hill, Harvard Graduate School of Education, 6 Appian Way, Cambridge, MA 02138, United States. Email: heather_hill@gse.harvard.edu

Structured Reporting Guidelines for Classroom Intervention Research

Abstract

Inconsistent reporting of critical facets of classroom interventions and their related impact evaluations hinders the field's ability to describe and synthesize the existing evidence base. In this essay, we present a set of reporting guidelines intended to steer authors of classroom intervention studies toward providing more systematic reporting of key intervention features and setting-level factors that may affect interventions' success. The guidelines were iteratively developed using recommendations and feedback from scholars active in conducting and synthesizing classroom intervention research. This effort aims to open wider the 'black box' in classroom research, communicating key information with more precision and detail to practitioners and future researchers, and permitting the field to more efficiently accumulate and synthesize findings on classroom interventions, determining what works, for whom, and under what conditions.

Keywords: impact evaluations; meta-analyses; classroom interventions

Structured Reporting Guidelines for Classroom Intervention Research

For decades, scholars have designed programs that aim to change both the content taught in classrooms and the instructional strategies used to teach that content. Many of these classroom intervention programs feature new curricula, teacher professional development, or both, and involve significant teacher learning. The push for randomized experimental research on classroom interventions (National Research Council, 2002) marked a critical juncture for the field, as a wave of RCTs brought insights about the impacts of a wide range of initiatives using designs supporting causal inference. Scholars have used these results to develop and adapt interventions (e.g., Jones et al., 2019; Kim et al., 2024; Westbrook et al., 2023) and to produce research syntheses focused on reading, mathematics, and science programs (e.g., Didion et al., 2020; Lynch et al., 2019; Taylor et al., 2018), classroom technology (e.g., Cheung & Slavin, 2012, 2013), and social-emotional learning (Cipriano et al., 2023), among other topics. One important goal for such syntheses is to describe programs and their effects. Another important goal of many of these reviews is explaining treatment heterogeneity – modeling program outcomes as a function of both program and setting-level characteristics, with the goal of identifying specific features that predict enhanced student learning.

A common problem for those attempting to survey and understand this evidence base involves the under-reporting of important programmatic and setting-level factors. Often, the discussion sections of research syntheses and reviews point out the vagueness in original study reports about what the evaluated programs actually entailed and the contexts in which they were enacted (see, e.g., discussions in Garrett et al., 2019; Kraft et al., 2018; Lynch et al., 2019; Scher & O'Reilly, 2009). These omissions limit readers' understanding of the conditions underlying original study impacts and restrict the conclusions that research synthesists can draw. For

example, although case study evidence backs the notion that the kinds of resources provided by the school or district to support programs is a contributing factor to their success (e.g., Weiss et al., 2014), few published impact reports contain information about such support. Also frequently missing is information about teacher attendance at and active participation in program activities. Finally, even some contemporary published reports do not contain descriptive information about the student sample (e.g., Jaciw et al., 2020), and most do not contain enough information to allow meta-analysts to test for differences in program impacts across key groups, such as students from underrepresented racial/ethnic backgrounds or students who are dual language learners.

The under-reporting described above could stem from several sources, including the financial, logistical, and measurement challenges associated with collecting such data; a lack of awareness that such factors may influence program outcomes; and norms in the field that encourage authors to report “what worked” without simultaneously exploring potential reasons why. Importantly, however, unlike standards written to systematize reporting about data collection and analyses (Appelbaum et al., 2018), no detailed reporting standards exist for describing classroom intervention program characteristics, setting-level characteristics, and specific group results in a way that would support readers’ understandings of the conditions underlying study impacts and enable later examinations of heterogeneity in effects.

Without this information, researchers’ ability to accumulate evidence from the classroom intervention research base is constrained in three key ways. First, missing information about key program and setting-level characteristics means the field cannot ascertain whether these features tend to predict improved classroom quality and student outcomes. This limits researchers’ ability to fully understand existing work, build theory, and iteratively design more efficacious

interventions (e.g., Century & Cassata, 2016). Second, missing information about program and setting-level characteristics means these become omitted variables in meta-analyses; reducing the possibility of such bias is desirable.

Third, although we began this work as meta-analysts with the goal of enhancing the accumulation of evidence about how, when and why programs work, we are also authors of classroom intervention studies. From this perspective, we argue that these guidelines can enhance reporting on interventions more generally, enabling authors to communicate key information with more precision and detail to policy-makers, practitioners and future researchers. Such information may help the field in three ways. First, it will help those seeking to replicate or build upon successful programs to know what program features and setting-level characteristics were present during the initial trial. These features and characteristics may be necessary for successful replication (e.g., Makel & Plucker, 2014). Second, this information may also help the field develop a common language for discussing instructional improvement efforts, one that facilitates precision in communication and allows more productive dialogue, particularly around the techniques used by programs to change the content and instructional strategies in classrooms. Finally and more practically, these guidelines can prompt authors to plan for and collect important data.

To this end, this paper proposes a set of guidelines, organized into six categories, for reporting on interventions intended to change or improve classroom instruction and to lead to better student learning. We recommend that authors, editors, and reviewers use these guidelines in conjunction with others, noted below, that support the reporting of research design and methodological information. We describe the development of these guidelines next.

Constructing the Guidelines

Background

The guidelines presented below sit among several parallel efforts intended to improve the completeness and clarity of reporting on evaluative studies. Perhaps the best known among these efforts are the Consolidated Standards of Reporting Trials (CONSORT) and the American Psychological Association's Journal Article Reporting Standards (APA-JARS), both of which provide authors with a checklist of research design and methodological features recommended for inclusion in published reports. Both briefly address the reporting of intervention content, with CONSORT advising that authors should describe "the interventions for each group with sufficient details to allow replication, including how and when they were actually administered" (Schulz et al., 2010, p. 2) and APA (2020) specifying that authors should report the content, method of delivery, provider, setting, duration, time span, and "sufficient detail to allow for replication" (p. 1). In education, the Standards for Excellence in Education Research, published by the U.S. Department of Education, include a requirement that "researchers ... document the components of an intervention, including its essential practices and structural elements" (downloaded 12/22/23). C. J. Hill and colleagues (2023) built upon this work in a research guide for studying educational program implementation, providing recommendations and considerations for reporting on programmatic and environmental features in impact evaluations conducted in a range of education and human service settings.

Although such guidance forms a useful basis for a description of interventions, we argue that understanding program impacts and accurately explaining treatment heterogeneity for classroom interventions hinges on reporting program features and setting-level characteristics with more specificity and uniformity than contained in these prior sets of reporting standards. Thus this current work builds on prior efforts by generating guidelines for reporting broadly on

interventions intended to change the content taught or pedagogical strategies used in classrooms. These reporting guidelines are tailored to these classroom interventions, reflecting both theoretical views of how classrooms change and the ways existing interventions have approached this task. In doing so, the current work aligns with initiatives like Scher and Martinez (2022), who proposed reporting guidelines for intervention components specific to programs in a well-defined space (postsecondary developmental mathematics and foundational literacy). Other fields have similarly taken up this reporting challenge by creating their own tailored guidelines. Several teams have made field-specific adaptations to CONSORT, including for social and psychological interventions (Montgomery et al., 2018), and the U.S. Department of Health and Human Services has made a reporting checklist available for programs that address teen pregnancy (HHS Office of Population Affairs, 2023).

Developing the Guidelines

We began by delineating the set of studies whose reporting we aim to improve via these guidelines. Specifically, the guidelines apply to studies of *classroom interventions*. We define classroom interventions as programs that seek to alter the content delivered to students and/or the pedagogy used to deliver that content; that target all students within a classroom with new instruction, rather than targeting only a subset of students or individual students (e.g., via tutoring or pull-out remediation); and that are implemented primarily by classroom teachers. Such programs may be provided by researchers working within a university, private firms, non-profits, or others who offer services with the intent of changing classroom instruction. We include programs in all disciplinary areas (e.g., mathematics, English Language Arts, social studies) as well as programs meant to change other features of classrooms, including student

behavior, student social-emotional learning, and learning technologies that help co-teach content with teachers (e.g., Pane et al., 2014).

To build the guidelines, we started with theoretical studies, including D. K. Cohen's (2011) description of resources for teaching and Grossman, Compton et al.'s (2009) description of teacher education pedagogies. We next started to list potential guidelines by reviewing existing program typologies (e.g., Kennedy, 2016) and descriptions of setting-level characteristics (e.g., Durlak & DuPre, 2008; C. J. Hill et al., 2023; Weiss et al., 2014) that can influence instruction and student outcomes. We expanded our list by examining meta-analyses and synthetic reviews of classroom intervention studies, seeking the program features and setting-level characteristics authors focused on (Corcoran et al., 2018; Desimone, 2009; Didion et al., 2020; Kennedy, 2016; Kraft et al., 2018; Lynch et al., 2019; Sims et al., 2023; Slavin & Lake, 2008; Slavin, Lake, Chambers, et al., 2009; Slavin, Lake, & Groff, 2009; Taylor et al., 2018; Timperley et al., 2007; Wilson & Berne, 1999). We also conducted a review of existing reporting guidelines, including those in JARS, CONSORT, and CONSORT-SPI. As we engaged in this work, we also convened a Society for Research on Educational Effectiveness (SREE) conference session (described below) to gather input.

Following these steps, we distilled the list of program features and setting-level characteristics into a tractable number by eliminating overlapping categories (e.g., "external expertise" and "school-based vs. external providers") and ambiguous or seldom-used categories (e.g., "prevailing discourse challenged"). We then organized our list using a logic model adapted from Scher and O'Reilly (2009) (see Figure 1). This logic model places classroom interventions at its center, then asks for detailed information about five dimensions along which classroom

interventions might vary in ways that impact program outcomes. The five dimensions are as follows:

1. ***Program provider and intervention content for students***; the latter refers to the materials or classroom instruction meant to be adopted by teachers and experienced by students.
2. ***Program content for teachers*** (e.g., bodies of knowledge; prescriptions for practice) through which programs attempt to influence classrooms.
3. Characteristics of any ***program professional development*** (e.g., duration, format and activities).
4. Features of ***program curriculum materials*** (e.g., student contact hours and time span; educative features for teachers).
5. ***Ongoing support for implementation*** (e.g., lesson materials, support meetings) that can create a sustaining environment for the intervention.

In line with the logic model, the guidelines also ask for information about the ***setting-level resources contexts, and barriers*** in which interventions are implemented, including both supports and barriers to implementation present in the school, district, and community environment in which programs are enacted, as well as the resources that teachers themselves bring to implementation. Finally, the guidelines ask for reporting of ***impacts by specific student groups***, which will support analysis of whether interventions have differential effects on relevant student populations.

The guidelines benefitted at each stage of their development from input from the classroom intervention research scholarly community. At an early stage, we convened a session at the annual SREE conference in order to brainstorm important features for inclusion in the

guidelines. We shared a prototype structure for the guidelines, along with potential categories and codes, then session participants worked in groups to identify content they considered important to report under each category. We used input from this session to inform the initial guidelines draft, e.g., including a section on ‘specific group reporting’ to enable future researchers to examine impacts on more targeted groups of students.

We then solicited close review and feedback of the draft guidelines from a total of 17 scholars, among them a number of *JREE* authors, who have, in recent years, a) designed interventions intended to improve classroom practice; b) conducted randomized impact evaluations; and/or c) conducted meta-analyses of classroom-level interventions. We began by asking four senior scholars in the fields of curriculum, professional development, special education, and language and literacy development, respectively, to review and provide initial feedback on the portions of the guidelines germane to their expertise. After we arrived at an initial draft of the guidelines, we conducted interviews with 10 scholars, asking them to think aloud as they described a specific program they created or evaluated in response to the items on our list. In this phase, we selected researchers to represent a range of content areas (e.g., mathematics, reading, social studies, SEL, computer science) and grade levels served (pre-K to adolescents), and to include researchers with expertise in both program design and evaluation. Most researchers had been involved in federally-funded research studies supported by NSF or IES. We used these interviews to check whether the wording of each specific guideline elicited the intended information, adding supportive information or changing the wording when it did not (e.g., “summer workshops” changed to “workshops” after a respondent noted that her program held workshops in the fall). Results from this feedback suggested that the guidelines as written generally elicited the intended descriptions. We also collected information about missing topics,

adding these to the list (e.g., collecting not only provider name but also provider's relationship to the evaluation). We concluded this review stage by obtaining further feedback on this manuscript and proposed guidelines from an additional set of three senior scholars with extensive experience in a) teacher professional development and curriculum theory and practice; and b) pre-K-12 educational impact evaluations.

We exclude two topics from these guidelines. First, authors should refer to APA-JARS as well as IES's What Works Clearinghouse reporting standards (WWC, 2021) when deciding how to report research design, methodological, and sample information, including participant population and location where the study occurred. Second, while we recommend that authors name program content (e.g., project-based learning in science), we do not create a detailed breakdown of the types of content that could be presented to teachers (see Scher & Martinez, 2022, for one effort to do so).

Finally, because the achieved program is most relevant to replication and meta-analysis, the guidelines call upon authors to report program features as implemented. However, when program elements (e.g., coaching, an online component of PD, follow-up) were planned but not implemented, authors should note this briefly. As is custom in most reports, authors should report classroom observation measures of fidelity of implementation as program outcomes.

Guidelines for Reporting on Classroom Interventions

Below, we present the guidelines' reporting categories, organizing them by program features, setting-level characteristics, and specific group reporting. We considered alternative organizational structures (e.g., based in behavioral economics) but settled on organizing by program features and setting-level characteristics because of their theoretical importance and prevalence as conceptual organizers in the field (i.e., C. J. Hill et al., 2023; Scher & O'Reilly,

2009). In the text and accompanying tables, we provide definitions for and examples of these categories. Importantly, the categories are not mutually exclusive, and information about the same program element may be reported in several different tables. For instance, a new curriculum may contain educative features, such as information meant to build teacher content knowledge. This would be coded under Table 4, ‘Program Curriculum Materials,’ row 4, ‘Educative Features.’ Simultaneously, this same curriculum-provided information meant to build content knowledge would also be coded under Table 2, ‘Program Content for Teachers,’ row 2, “Bodies of Knowledge.”

We also make recommendations for how to report information within these categories—for example, descriptively, numerically, categorically, or with Likert-type descriptors. However, we encourage authors to embed this information in narrative descriptions of the program, providing as much detail as space allows. In other words, authors should continue to provide rich descriptive information about their program, similar to that found in Carpenter et al. (1989), Garet et al. (2016), and Taylor et al. (2017). Appendix A provides a written example of integrating the reporting guidelines with descriptive information about a program, using an existing published report to do so.

In Online Appendix A, we illustrate how eight diverse programs and program evaluations could be described using the reporting guidelines. Again, we do not suggest authors should report intervention characteristics in this format, preferring the elaborated descriptions modeled in Appendix A. Instead, our goal was to give authors an opportunity to learn how to use the guidelines by providing our sense of how reporting for these well-known programs would look. We did not code for basic information about the program provider and intervention content for

students, as this information is nearly universally already provided in research studies, or for setting, because that information is largely unavailable in the published reports examined.

Program Provider and Intervention Content for Students

Table 1 displays the reporting guidelines for *Program Provider and Intervention Content for Students*. We begin by noting that authors should name the provider of the classroom intervention, describe the provider's sector, and explain the provider's relationship with the evaluation. Authors should also describe intervention content for students – what topics students would be learning, and the pedagogy used to teach them – at a level detailed enough to enable future readers and/or research syntheses to generate meaning from such information. Authors should report, to the extent possible, instructional improvement content experienced by the control group.

Program Content for Teachers

Table 2 presents the reporting guidelines for *Program Content for Teachers*. These comprise three classes of program content (i.e., what is taught to teachers) that reflect three different viewpoints about how to most effectively strengthen classroom instruction. Two categories provide tools for teachers to use when designing instruction or making decisions in the moment; the third is defined by more prescription of classroom activities. Authors should describe the content as implemented by their program during the study.

Conceptual Frameworks

Programs may contain “principles, frameworks, or guidelines that teachers use to guide their decisions about teaching and learning” (Grossman, Hammerness, & McDonald, 2009, p. 274). Following Grossman, Hammerness, and McDonald (2009), these frameworks can include “general, applicable theories such as constructivist theories of learning, motivation, and

instructional scaffolding or more philosophical views related to the purposes of schooling such as social justice and the goal of improving educational opportunities for historically under-served students” (p. 274). While conceptual frameworks can guide teachers’ thinking about and design of instruction, they do not offer concrete advice about specific classroom routines, instructional strategies, or student tasks. Instead, teachers are meant to infer specific actions from the broader conceptual frameworks presented by the program.

Bodies of Knowledge

Programs may attempt to influence classrooms by helping teachers learn a body of knowledge (Kennedy, 2016). To differentiate from conceptual frameworks, we define bodies of knowledge as knowledge based in subject matter discipline (e.g., fractions) or empirical study (e.g., how students conceptualize fractions). Like conceptual frameworks, bodies of knowledge do not prescribe teacher actions—they are meant to shape teachers’ thinking about and design of instruction and influence decision-making in the classroom.

Prescriptions for Practice

Programs may also offer prescriptions for practice (Kennedy, 2016), meaning resources that outline and support specific interactions between teachers and students. Prescriptions differ from both conceptual frameworks and bodies of knowledge in that they provide concrete activities for teachers or students to engage in. As Kennedy (2016) describes, “prescriptions reduce the amount of individual discretion or judgment that is needed, thus ensuring that teachers do things exactly as the provider intends” (p. 11). Prescriptions for practice are embedded, for instance, in lesson materials (including curricula) and structured classroom routines (e.g., classroom management routines meant to promote student self-regulation; math talks in mathematics education).

Treatment Contrast

To the extent possible, authors should report on the PD program content experienced by the control group.

Program Professional Development

Table 3 displays the reporting guidelines for *Program Professional Development (PD)*. Program designers make decisions that shape the time and space in which teachers learn new content while in professional development settings; they also help determine PD program content, or arrange to share that decision with district or school personnel. These decisions often connect to larger views of how to best change instruction—for example, that teachers need time, job-embedded learning opportunities, collective participation with peers, or to provide input that shapes the professional development agenda. Authors should report these topics, in part because they are frequently raised in meta-analyses (e.g., Didion et al., 2020) and other research on teacher PD (e.g., Desimone, 2009), indicating scholars and program designers see them as consequential to teacher learning. Authors should describe the characteristics of program PD as implemented by their program during the study, and briefly report PD that was planned but not implemented.

Overall Duration

Authors should provide the number of PD contact hours delivered during the study and the maximum time span over which teachers could participate in the PD. Teacher attendance at PD activities is discussed in the *Teacher Participation* portion of these guidelines.

PD Formats

Professional development programs use a variety of formats to deliver new content to teachers. Desimone (2009), Borko and colleagues (2010), and others describe common formats

and authors should report them here, including summer or school-year workshops; coaching; professional learning communities; the collective participation of teams of teachers from given school sites; and whether any aspect of the PD was online.

Who or What Determines PD Content

Authors should report who decides the content of PD. Most typically in classroom intervention research, program designers choose content themselves, and thus this is the default category if no other information is reported. However, programs may also invite teachers' input into PD content—for instance, allowing teachers to choose topics or co-create content with PD providers—potentially resulting in a better fit of PD to teachers' needs and increasing teachers' engagement (H. Hill et al., 2021). Inviting school leaders to choose or co-create PD content may similarly result in a better fit of the program content to school needs (e.g., Stevenson et al., 2016). Finally, analyzing school or district academic and/or student engagement data to determine gaps and growth needs may also result in an improved fit between the PD and local contexts (Tallerico, 2005).

PD Activities

Similar to K-12 education, where teachers convey content to students via specific pedagogical methods, PD interventions meant to improve classroom instruction also convey program content to teachers via specific pedagogical activities (see Grossman, Compton et al., 2009; Kennedy, 2016). As we describe below, these pedagogical activities are often based on implicit or explicit theories about how to change classrooms. The guidelines include six pedagogical activities by which PD programs often aim to teach educators the PD content. To the extent programs use additional types of pedagogical activities, authors should describe these as well.

Representations and/or Decompositions of Practice. Programs may provide representations, or demonstrations, of instructional practice via video, live models of teaching, case studies, and so forth to build teacher knowledge of a specific practice. Programs may further help teachers decompose a demonstrated practice, breaking it down to analyze its constituent parts (Grossman, 2011).

Approximations of Practice. Programs may offer teachers opportunities to engage in deliberate practice of new instructional strategies and routines outside of their actual classroom, often via rehearsals (Lampert et al., 2013) or via technology-enhanced teaching simulations (J. Cohen et al., 2020). Typically, a facilitator, peer, or coach provides feedback on teachers' practice with the new skill; other educators may also participate by acting as "students" to simulate classroom conditions. By using approximations of practice rather than "live" practice in an actual classroom, programs may help teachers refine and build their teaching in a controlled environment (Lampert et al., 2013).

Teacher Reflection. Programs may also prompt teachers to engage in reflection on their practice (Schön, 1983). Reflection activities may help teachers gain insight (Kennedy, 2016) into their existing practice, new instructional practices, student responses to instruction, or any number of other topics related to classrooms. Reflections on practice can occur individually, often in comparison to an idealized version of instruction (Hatton & Smith, 1995) or in the context of conversations with coaches or peers. Reflection on practice is often meant to shape and improve in-the-moment instructional decision-making (Schön, 1983).

Analysis of Student Work and/or Performance. Programs may present two kinds of evidence about students intended to improve instruction. First, programs may present student work – e.g., videos of students solving problems, written essays – and ask teachers to study and

learn from this work. Programs may also present teachers with student assessment data and then support them to analyze those data for the purpose of instructional improvement (e.g., Goertz et al., 2009; West et al., 2016). Such data can take the form of student performance on state assessments, interim assessments supplied by commercial vendors, formative assessments supplied by the program, or classroom assessments typically used by teachers.

Planning Instruction. Programs may ask teachers to plan instruction (e.g., lesson plans, instructional tasks) using the information and materials provided by the program. Allowing time and scaffolding teacher planning may assist teachers in integrating the lessons from the program into their classroom practice (e.g., Carpenter et al., 1989; Penuel et al., 2007).

Assistance in Learning to Implement Curriculum Materials. Finally, programs may help teachers learn about and implement curriculum materials. For instance, programs may ask teachers to analyze and ‘internalize’ high-quality curriculum lessons or units (Short & Hirsh, 2022). In these sessions, teachers may complete sections of the materials as if they were students, solving problems or conducting inquiry to help familiarize themselves with the disciplinary content and instructional activities contained in them (e.g., Cohen & Hill, 2001). An explicit focus on curriculum materials may help teachers translate learning from PD more efficiently to their classroom practices and may improve the quality of curriculum implementation (Cohen & Hill, 2001).

Teacher Participation

Teachers’ attendance at program events and activities, as well as their engagement and active participation while present at program events and activities, shape teachers’ opportunities to learn (H. C. Hill & Erickson, 2018). Teacher attendance and engagement may be influenced by whether program providers and/or the schools and districts in which they operate require

teacher participation, or whether teachers voluntarily enroll in the program. Programs for which teacher attendance is mandatory may see smaller impacts compared to those in which teachers volunteered to participate (e.g., Kennedy, 2016).

Treatment Contrast

Authors should report on treatment contrast by comparing PD experiences for treatment and control group teachers. This may take the form of collecting information about PD duration, content, format, activities, and participation structures for both groups and either providing a direct comparison or summarizing the PD experienced by control group teachers.

Program Curriculum Materials

Table 4 presents the guidelines for *Program Curriculum Materials*. Like professional development, curriculum materials can shape classroom instruction in significant ways, including through providing prompts, tasks, and activities for teachers to use with students, and through any teacher-focused opportunities to learn embedded in those materials (Ball & Cohen, 1996). Here, we list characteristics of student-facing materials—whether disciplinary (e.g., mathematics textbooks) or content-general (e.g., materials that support SEL programs)—that authors should report if implemented in their study.

Curriculum Contact Hours and Time Span

Curriculum materials vary in their intensity, or “dosage,” to which students are intended to be exposed, and teachers’ take-up of curriculum materials also varies across interventions. Authors are asked to report both the proportion of a typical lesson for which the new curriculum materials were intended by the program developers to replace existing materials, and an estimate of the percentage of the intended curriculum materials that were actually used by teachers in the study. Authors should also report the duration (span) of time over which the program developers

intended for the new curriculum materials to be used, as well as an estimate of the actual mean span of time over which the materials were utilized by teachers in the study. These indicators of intensity can affect outcomes collected on a yearly basis, such as student state test scores (see, e.g., Roblin et al., 2018; Stockard et al., 2018).

Online Content

Curriculum-focused programs can vary in the extent to which students receive instruction from classroom teachers versus from online resources (e.g., animated mathematics lessons). Although the guidelines exclude programs that deliver content solely online, mixed programs that intend to change teachers' instruction do exist (e.g., Pane et al., 2014) and such modalities should be described.

Educative Features

Curriculum materials can be educative for teachers, meaning they intentionally provide information meant to change teachers' knowledge, beliefs, and practice (Davis & Krajcik, 2005; Roblin et al., 2018). Materials might, in their design, guide teachers' attention to the "big ideas" in lessons or provide ready-to-use, student-friendly definitions. They may educate teachers via narratives or case studies that describe high-quality enactment of a lesson or by presenting information about student learning trajectories (Clements & Sarama, 2008; Davis et al., 2017). Finally, materials may contain "design rationales" that explain lesson or unit design and sequencing, and how these lessons and units connect to the broad goals of the curriculum (Davis & Krajcik, 2005; Remillard & Kim, 2020; Remillard et al., 2019).

Suggestions for Adaptation

Curriculum materials may include explicit suggestions for omitting or adjusting their content, for example in response to perceived student needs and/or time constraints (Ball &

Cohen, 1996; Davis et al., 2017). Such guidance may not only help teachers adapt the materials more efficiently, but also help the adaptations to remain in line with the original program intent.

Treatment Contrast

Because using similar materials in the treatment and control condition may attenuate impacts (Weiss et al., 2014), authors should report the materials used in the control condition. And, because not all features of curriculum materials will be captured in the indicators above, to the extent feasible given the constraints of proprietary materials, we recommend making a sample lesson or unit from the treatment and comparison/control teachers' editions available in an online appendix. This will allow future meta-analysts and other researchers to code for features of materials that may be consequential to student learning—for example, particular ways of representing content or specific instructional routines— that cannot be anticipated at this time.

Program Support for Implementation

Table 5 shows the guidelines for *Program Support for Implementation*. Ongoing support may bolster educational interventions' traction in the classroom, making them more likely to be taken up and used over an extended period. Building forms of continuous assistance into instructional interventions may foster “sustaining environments” (Bailey et al., 2020) by providing continuity in educational support for teachers (Ramey & Ramey, 2007). Authors should report on the presence of two types of ongoing supports as implemented by their program during the study.

Student-Facing Materials

Programs may provide student-facing materials that help teachers enact program-intended instruction in their classrooms. These materials might include prompts, tasks, curricula, kits, formative assessment items, lesson plans, or other artifacts aligned to the program and

meant to be used by students. Recent meta-analyses note that some such materials can enhance the efficacy of STEM teacher professional development (Lynch et al., 2019) and coaching (Kraft et al., 2018).

Implementation Support Meetings

Many programs provide follow-up by hosting implementation support meetings after the initial phase of classroom use of the intervention. In contrast to the formats listed in Table 3 (e.g., coaching, workshops), these meetings are not meant to share new program content, but instead are meant to serve to fine-tune implementation, opening space for questions and troubleshooting, minor adjustments based on context, and teacher-to-teacher collaborative adaptation. A recent meta-analysis found such meetings enhanced the efficacy of STEM teacher professional development (Lynch et al., 2019).

Setting-Level Resources, Contexts, and Barriers

Table 6 displays the reporting guidelines for *Setting-Level Resources, Contexts, and Barriers*. Compared to intervention features displayed in Tables 1-5, for which the empirical and/or theoretical research bases supporting hypothesized links to intervention effectiveness are more well-developed, comparatively less research has investigated the impacts of setting-level factors on the outcomes of classroom-level interventions. Prior research has posited that classroom interventions are nested within the ecologies of schools, districts, and communities which differ in their resources, personnel, organizational environments, and local priorities (Scher & O'Reilly, 2009; Wilson, 2013), and hence their capacities to support instructional innovations. As such, scholars have considered how features of the context in which an intervention takes place may influence program implementation and outcomes (e.g., Coburn & Russell, 2008; Matsumara et al., 2009; Weiss et al., 2014; Wilson, 2013).

Because the number of contextual factors that could hypothetically influence the outcomes of classroom-level interventions is extremely large, specifying a parsimonious set of reporting guidelines for which study authors can pragmatically collect and report data is challenging. Nevertheless, we argue that the current lack of reporting guidelines has contributed to a condition where very little information on study setting-level supports is reported at all (Kraft et al., 2018; Lynch et al, 2019; Wilson, 2014). As such, putting forth reporting guidelines for contextual factors is an important step forward for the field, as incremental progress toward more consistent reporting on settings could advance the field's cumulative understanding of what works under what conditions.

In developing these recommendations, we aimed to balance prioritizing key constructs for reporting with consideration for the practical resource constraints on data collection that researchers routinely face. This work builds upon Hill et al. (2023), which posited several broad sets of contextual factors that could affect implementation of a range of educational interventions. The current guidelines are more specifically grounded in major findings intersecting the literatures on preK-12 classroom-level interventions and implementation (e.g., Coburn & Russell, 2008; Durlak & DuPre, 2008; Valli & Buese, 2007; Weiss et al., 2014), and hence cover a more targeted set of categories specific to the classroom intervention context. Similar to Hill et al. (2023), study authors should combine the context guidelines with their own knowledge of the research base on related prior interventions, in order to develop, collect data for, and report on a concise set of hypotheses about contextual factors likely to influence implementation of their specific program.

The contextual reporting guidelines are divided into three categories. Similar to *Program PD*, authors can state that elements not specifically named in their description of settings and

contexts were not present and/or did not affect program implementation.

District/School/Community Resources

Resources Provided by the District. The level and kinds of resources provided by the district may affect schools' technical capacities and readiness to implement new instructional programs (e.g., Waters & Marzano, 2006; Weiss et al., 2014). District staff may support implementation via mechanisms such as financial allotments for curriculum materials or program staff, encouraging principals or schools to adopt and support a program, providing additional time for teacher professional development, assigning a district staff member to be actively involved in implementation, or by providing material and technological support for the program, such as meeting space. District officials may also select and support interventions that align with their strategic improvement plans, creating incentives for schools and teachers to adopt the program. In addition to supporting program implementation, district staff may encourage participation in a research study evaluating the efficacy of a classroom intervention, for instance by aiding in recruitment or providing teachers with professional development credits for participation.

Resources Provided by the School. As with districts, schools may provide (or not provide) supports that can influence intervention implementation, such as funds for materials and staff, protected time for professional development, and other material and technological support.

Resources Provided by the Community. Members of the communities in which schools are embedded may also provide resources that influence program enactment. Different from school- or district-provided resources like money and staff, these can take the form of community expertise as well as family and community funds of knowledge (González et al., 2006) used to shape intervention content and design (see, e.g., Kisker et al., 2012; McWayne et

al., 2021). These may also include financial and in-kind support from community institutions such as museums, cultural organizations, etc.

Contexts for Implementation

School Leader Messaging. Individuals in school leadership positions may bolster program implementation by publicly endorsing an intervention, encouraging teachers to participate in it, serving as ‘program champions’ who clear bureaucratic roadblocks for a program, or engaging in program activities and events themselves (see, e.g., Durlak & DuPre, 2008; Kearns et al., 2010; Valli & Buese, 2007). Leaders can explain how the intervention fits into a coherent overall blueprint for school improvement and ensure teachers feel safe during periods of experimentation and change. Individuals in leadership positions can also get in the way of programs by failing to endorse or support them, or by disseminating messages to teachers that conflict with the goals of the intervention (see, e.g., Coburn & Russell, 2008).

School Climate. Prior research has suggested that instructional interventions may be more likely to succeed when they are implemented in schools that feature a positive school climate, where teachers feel respected, experience collaborative relationships with colleagues, and have reasonable workload demands (Johnson et al., 2012; Sebring & Bryk, 2000).

Teacher Resources. Teachers’ own levels of support for a new program, as well as their prior knowledge, self-efficacy, affect, and beliefs relevant to the intervention, can all affect their level and quality of program take-up (Durlak & DuPre, 2008; Hill & Erickson, 2019; Kearns et al., 2010). Researchers should consult the extant literature base on interventions with overlaps to their own in order to gauge which teacher-level characteristics are most likely to influence implementation, if any (C. J. Hill et al., 2023), and report on these characteristics as resources permit.

Barriers

Competing Priorities or Initiatives. The presence of competing school or district initiatives may constrain interventions' effectiveness if teachers prioritize these over the programs being evaluated (Coburn & Russell, 2008; Hill et al., 2018).

Turnover of School and/or District Leaders. Turnover of district or school leaders, particularly those individuals who initially supported the intervention, can threaten its continued implementation (Hill et al., 2018; Weiss et al., 2014).

Lastly, while not an explicit reporting category, researchers should note any unexpected shocks or salient political or environmental factors that could be expected to appreciably interact with the specific intervention, affecting implementation (Durlak & DuPre, 2008; C. J. Hill et al., 2023). For example, if a civics curriculum intervention were enacted in a highly politically divided school district amidst a contentious presidential election, this context could be expected to affect how the intervention was enacted and would warrant reporting.

Treatment Contrast

Typically, control group teachers and schools do not receive intervention-specific resources or messaging. Further, in experimental studies, treatment and control conditions in place at baseline related to teacher resources and school climate should be equal in expectation due to random assignment. Nonetheless, group imbalance is always possible, particularly in studies with small sample sizes. Further, for quasi-experimental studies, baseline treatment/comparison group differences in these setting features may occur. Overall, authors should report any relevant differences between treatment and control groups that are predicted to affect the study outcomes.

Specific Group Outcomes

Because interventions can differ in their effectiveness for different populations, authors should present key program impact estimates broken out by the specific student groups served by the study. Doing so allows meta-analysts the opportunity to conduct more nuanced analyses of *for whom* programs work. The current practice of reporting overall summary statistics and overall sample characteristics allows meta-analysts to observe the association between the size of treatment effects and the proportion of students in a study who are members of a specific group. However, meta-analysts may also wish to investigate whether the effects of programs or programs with a specific feature tend to differ for children from different sociodemographic groups. To do so, meta-analysts must have access to information on the impacts of a treatment for each specific group within each study.

For this reason, we recommend that study authors report impact data (e.g., means and standard deviations of the outcome measures) separately for children within specific subgroups of theoretical and policy relevance to the study setting and context, ideally after specifying this planned reporting in a pre-analysis plan. To do so, authors should report sample characteristics and treatment impacts in ways consistent with the APA's recently released JARS-Race, Ethnicity, and Culture (JARS-REC) standards. Table 7 demonstrates how a primary study might present this information. We also recommend that study authors collect and report baseline data for the overall sample and subgroups when possible, consistent with the What Works Clearinghouse (U.S. Department of Education et al., 2022) recommendations.

Discussion and Conclusion

This paper proposes a set of guidelines for describing classroom interventions. These guidelines are meant to be used in conjunction with APA-JARS and IES-WWC, which provide detailed guidance regarding how to report study research design and methods. If used widely,

these guidelines would support those who wish to build from, adapt, or adopt these interventions by elaborating in more detail on key program features and contextual conditions supporting implementation and student success. They would also provide authors of research syntheses with critical information for testing hypotheses about effective program features, how setting-level variables predict outcomes, and how program impacts differ by the specific groups served. As the number of impact studies of classroom interventions continues to grow, ensuring the ability to test hypotheses in these areas expands the reach of evaluation research beyond simple summary evaluations of impact.

These guidelines face several challenges, ones that the field must solve together. One is adoption. Study authors can play a key role in building evidence for the field by using these guidelines based on their value for communicating key information and for supporting later meta-analysis. Presenting important information in a standardized format may also increase the discoverability and comprehensibility of study authors' work, increasing the chances that study reports are both cited and built upon in future work. To make the guidelines more user-friendly for authors, the supplemental appendix¹ includes a downloadable checklist that authors can use as a template to record information. Meanwhile, journals themselves can encourage the reporting guidelines in their "author instructions," and professional associations, study registries, clearinghouses, and federal funding agencies can suggest or require the use of these alongside the APA and IES-WWC guidelines. Because the APA and IES-WWC guidelines well-address study methodological features, but not features of the interventions themselves, the current guidelines uniquely contribute guidance for reporting on core features of classroom interventions.

¹ Note to reviewers: This checklist is forthcoming; we commit to producing it with the engagement of a graphic designer upon final approval of the guidelines for publication.

Another challenge is around the length of such descriptions in written publications, which typically have strict word limits. Appendix A shows that the addition of guideline information to an existing report generated around 400 words (shown in bold), providing an existence proof for the notion that this task can be manageable. As Appendix A shows, authors can save space by including a sentence indicating that program features not explicitly mentioned were not contained in the program, and that setting-level contexts not mentioned either did not exist or did not affect the implementation of the program. Authors may also, for the sake of completeness, seek to include some of this information in an online appendix. A related challenge is the additional data collection necessary to collect information, particularly on teacher attitudes, beliefs, attendance, and engagement; control group experiences; and setting-level supports and barriers. Teacher surveys – common to many evaluations but also expensive to implement – are necessary for some of these data collection efforts. Funding agencies, including NSF and IES, must take this into account when setting award levels and reporting guidelines.

These guidelines' limitations also point toward fruitful future directions for their continuous development and refinement. While this tool includes guidelines gleaned from an extensive review of the literature and feedback from scholars conducting classroom intervention research in the field, we could not include every construct. The guidelines also prompt study authors for qualitative elaborations/descriptions to explain their codes, permitting researchers to add important elaborative information that falls outside the existing categories.

Similarly, the guidelines will eventually become outdated as the field develops new strategies for fostering change in classrooms, such as by building new technological supports (Demszky et al., 2023). Similar to how other fields update their reporting standards (e.g, CONSORT), we intend that the guidelines will be incrementally refined via an iterative update

procedure. We will gather feedback from the inceptive wave of field implementation via surveys of researchers and an update-focused consensus meeting convening researchers and representatives from major professional societies, journals, and funders; assess new directions in instructional improvement (e.g., A.I.-driven teacher PD); then update materials on an open access website or publish V2.0 of these in ~5-10 years. Via these mechanisms, future researchers could add to the tool as knowledge in the field about key variables influential to classroom intervention impacts evolves.

Overall, these guidelines represent an improvement over current reporting standards, which lack detail and result in idiosyncratic and sporadic reporting of key features. By taking the next step as a field toward more consistent reporting, we can collectively learn more from our investments in classroom intervention research, ultimately contributing to stronger learning opportunities for students.

References

- American Psychological Association. (2020). *JARS-Quant: Table 2: General principles*.
<https://apastyle.apa.org/jars/quant-table-2.pdf>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology. *American Psychologist*, 73(1), 3–25.
- Bailey, D. H., Jenkins, J. M., & Alvarez-Vargas, D. (2020). Complementarities between early educational intervention and later educational quality? A systematic review of the sustaining environments hypothesis. *Developmental Review*, 56, 100910.
- Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, 25(9), 6-14.
- Borko, H., Jacobs, J., & Koellner, K. (2010). Contemporary approaches to teacher professional development. *International Encyclopedia of Education*, 7(2), 548-556.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.
- Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26(4), 499-531.
- Century, J., & Cassata, A. (2016). Implementation research: Finding common ground on what, how, why, where, and who. *Review of research in education*, 40(1), 169-215.
- Cheung, A. C., & Slavin, R. E. (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review*, 7(3),

198-215.

Cheung, A. C., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis.

Educational Research Review, 9, 88-113.

Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J.

(2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development*, 94(5), 1181-1204.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, 45(2), 443-

494.

Coburn, C., & Russell, J. (2008). Getting the most out of professional learning communities and coaching: Promoting interactions that support instructional improvement. *Learning*

Policy Brief, 1(3), 1-5.

Cohen, D. K. (2011). *Teaching and its predicaments*. Harvard University Press.

Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. Yale University Press.

Cohen, J., Wong, V., Krishnamachari, A., & Berlin, R. (2020). Teacher coaching in a simulated environment. *Educational Evaluation and Policy Analysis*, 42(2), 208-231.

Corcoran, R. P., Cheung, A. C., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research*

Review, 25, 56-72.

- 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.
- Davis, E. A., & Krajcik, J. S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3-14.
- Davis, E. A., Palincsar, A. S., Smith, P. S., Arias, A. M., & Kademian, S. M. (2017). Educative curriculum materials: Uptake, impact, and implications for research and design. *Educational Researcher, 46*(6), 293-304.
- Debarger, A. H., Penuel, W. R., Moorthy, S., Beauvineau, Y., Kennedy, C. A., & Boscardin, C. K. (2017). Investigating purposeful science curriculum adaptation as a strategy to improve teaching and learning. *Science Education, 101*(1), 66-98.
- Demszky, D., Liu, J., Hill, H. C., Jurafsky, D., & Piech, C. (2023). Can automated feedback improve teachers' uptake of student ideas? Evidence from a randomized controlled trial in a large-scale online course. *Educational Evaluation & Policy Analysis, 01623737231169270*.
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181-199.
- Didion, L., Toste, J. R., & Filderman, M. J. (2020). Teacher professional development and student reading achievement: A meta-analytic review of the effects. *Journal of Research on Educational Effectiveness, 13*(1), 29-66.
- 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, 327-350.
- Garet, M. S., Heppen, J. B., Walters, K., Parkinson, J., Smith, T. M., Song, M., ... & Borman, G. D. (2016). Focusing on Mathematical Knowledge: The Impact of Content-Intensive coburnTeacher Professional Development. NCEE 2016-4010. *National Center for Education Evaluation and Regional Assistance*.
- Garrett, R., Citkowicz, M., & Williams, R. (2019). How responsive is a teacher's classroom practice to intervention? A meta-analysis of randomized field studies. *Review of Research in Education*, 43(1), 106-137.
- Goertz, M. E., Oláh, L. N., & Riggan, M. (2009). *From testing to teaching: The use of interim assessments in classroom instruction*. Consortium for Policy Research in Education.
- González, N., Moll, L. C., & Amanti, C. (Eds.). (2006). *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. Routledge.
- Grossman, P. (2011). Framework for teaching practice: A brief history of an idea. *Teachers College Record*, 113(12), 2836-2843.
- Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, 111(9), 2055-2100.
- Grossman, P., Hammerness, K., & McDonald, M. (2009). Redefining teaching, re-imagining teacher education. *Teachers and Teaching: Theory and Practice*, 15(2), 273-289.
- Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, 11(1), 33-49.
- Hill, C. J., Scher, L., Haimson, J., & Granito, K. (2023). *Conducting implementation research in impact studies of education interventions: A guide for researchers* (NCEE 2023-005).

- U.S. Department of Education. <https://ies.ed.gov/ncee/pubs/2023005/>
- Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting program impacts: A brief review. *Educational Researcher*, 48(9), 590-598.
- Hill, H., Papay, J., Schwartz, N., Johnson, S., Freitag, E., Donohue, K., Berry, R., III, Loeb, S., Anderson, M., Baker, M., Cato Czupryk, B., Coddington, C., Ehlman, K., Maus, A., Meili, L., Paek, P., Romansky, C., Taylor-Perryman, R., Vu, H., ... Williamson-Zerwic, B. (2021). *A learning agenda for improving teacher professional learning at scale*. Research Partnership for Professional Learning.
- Jaciw, A. P., Nguyen, T., Lin, L., Zacamy, J. L., Kwong, C., & Lau, S. S. (2020). Final Report of the i3 Impact Study of Making Sense of SCIENCE, 2016-17 through 2017-18. Grantee Submission.
- Johnson, S. M., Kraft, M. A., & Papay, J. P. (2012). How context matters in high-need schools: The effects of teachers' working conditions on their professional satisfaction and their students' achievement. *Teachers College Record*, 114(10), 1-39.
- Jones, S. M., LaRusso, M., Kim, J., Yeon Kim, H., Selman, R., Uccelli, P., ... & Snow, C. (2019). Experimental effects of Word Generation on vocabulary, academic language, perspective taking, and reading comprehension in high-poverty schools. *Journal of Research on Educational Effectiveness*, 12(3), 448-483.
- Kearns, D. M., Fuchs, D., McMaster, K. L., Sáenz, L., Fuchs, L. S., Yen, L., Meyers, C., Stein, M., Compton, D., Berends, M., & Smith, T. M. (2010). Factors contributing to teachers' sustained use of kindergarten peer-assisted learning strategies. *Journal of Research on Educational Effectiveness*, 3(4), 315-342.
- Kennedy, M. M. (2016). How does professional development improve teaching?. *Review of*

Educational Research, 86(4), 945-980.

Kim, J. S., Burkhauser, M. B., Quinn, D. M., Guryan, J., Kingston, H. C., & Aleman, K. (2017).

Effectiveness of structured teacher adaptations to an evidence-based summer literacy program. *Reading Research Quarterly*, 52(4), 443-468.

Kim, J. S., Gilbert, J. B., Relyea, J. E., Rich, P., Scherer, E., Burkhauser, M. A., & Tvedt, J. N.

(2024). Time to transfer: Long-term effects of a sustained and spiraled content literacy intervention in the elementary grades. *Developmental Psychology*.

Kim, J. S., & Mosher, D. M. (2023). Structuring adaptations for scaling up evidence-based

literacy interventions. In S. Q. Cabell, S. B. Neuman, & N. P. Terry (Eds.), *Handbook on the science of early literacy* (pp. 253-268). Guilford Publications.

Kisker, E. E., Lipka, J., Adams, B. L., Rickard, A., Andrew-Ihrke, D., Yanez, E. E., & Millard,

A. (2012). The potential of a culturally based supplemental mathematics curriculum to improve the mathematics performance of Alaska Native and other students. *Journal for Research in Mathematics Education*, 43(1), 75-113.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and

achievement: A meta-analysis of the causal evidence. *Review of Educational Research*, 88(4), 547-588.

Lampert, M., Franke, M. L., Kazemi, E., Ghouseini, H., Turrou, A. C., Beasley, H., Cunard, A.,

& Crowe, K. (2013). Keeping it complex: Using rehearsals to support novice teacher learning of ambitious teaching. *Journal of Teacher Education*, 64(3), 226-243.

Lynch, K., Hill, H. C., Gonzalez, K. E., & Pollard, C. (2019). Strengthening the research base

that informs STEM instructional improvement efforts: A meta-analysis. *Educational Evaluation and Policy Analysis*, 41(3), 260-293.

- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304-316.
- Matsumura, L. C., Sartoris, M., Bickel, D. D., & Garnier, H. E. (2009). Leadership for literacy coaching: The principal's role in launching a new coaching program. *Educational Administration Quarterly*, 45(5), 655-693.
- McMaster, K. L., Jung, P., Brandes, D., Pinto, V., Fuchs, D., Kearns, D., Lemons, C., Saenz, L., & Yen, L. (2014). Customizing a research-based reading practice: Balancing the importance of implementation fidelity with professional judgment. *The Reading Teacher*, 68(3), 173-183.
- McWayne, C. M., Greenfield, D., Zan, B., Mistry, J., & Ochoa, W. (2021). A comprehensive professional development approach for supporting science, technology, and engineering curriculum in preschool: Connecting contexts for dual language learners. In *Supporting children's well-being during early childhood transition to school* (pp. 222-253). IGI Global.
- Montgomery, P., Grant, S., Mayo-Wilson, E., Macdonald, G., Michie, S., Hopewell, S., & Moher, D. (2018). Reporting randomised trials of social and psychological interventions: The CONSORT-SPI 2018 Extension. *Trials*, 19(1), 1-14.
- National Research Council. (2002). *Scientific research in education. Committee on Scientific Principles for Education Research*. Shavelson, R.J., and Towne, L., Editors. Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press
- Neuman, S. B., Samudra, P., & Danielson, K. (2021). Effectiveness of scaling up a vocabulary intervention for low-income children, pre-k through first grade. *Elementary School*

- Journal*, 121(3), 385–409.
- 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.
- Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, 36(2), 127-144.
- Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, 44(4), 921-958.
- Ramey, S. L., & Ramey, C. T. (2007). Early educational interventions: Principles of effective and sustained benefits from targeted early education programs. *Handbook of Early Literacy Research*, 2, 445.
- Remillard, J. T., & Kim, O.-K. (2020). Beyond the script: How curriculum authors communicate with teachers as curriculum enactors. In J. T. Remillard & O.-K. Kim (Eds.), *Elementary mathematics curriculum materials: Designs for student learning and teacher enactment* (pp. 141-160). Springer Nature.
- Remillard, J. T., Reinke, L. T., & Kapoor, R. (2019). What is the point? Examining how curriculum materials articulate mathematical goals and how teachers steer instruction. *International Journal of Educational Research*, 93, 101-117.
- Ridge, B. L., & Lavigne, A. L. (2020). Improving instructional practice through peer observation and feedback: A review of the literature. *Education Policy Analysis Archives*, 28, 61-61.
- Roblin, N. P., Schunn, C., & McKenney, S. (2018). What are critical features of science curriculum materials that impact student and teacher outcomes? *Science Education*,

102(2), 260-282.

Scher, L., & Martinez, A. (2022). *Developing a core components nomenclature in education: An update on IES-funded work*. Institute of Education Sciences. Retrieved December 22, 2023, from https://ies.ed.gov/seer/core_components.asp

Scher, L., & O'Reilly, F. (2009). Professional development for K–12 math and science teachers: What do we really know? *Journal of Research on Educational Effectiveness*, 2(3), 209-249.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. London: Temple Smith.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *Journal of Pharmacology and Pharmacotherapeutics*, 1(2), 100-107.

Sebring, P. B., & Bryk, A. S. (2000). School leadership and the bottom line in Chicago. *Phi Delta Kappan*, 81(6), 440-443.

Short, J. B., & Hirsh, S. (2022). *Transforming teaching through curriculum-based professional learning: The elements*. Corwin Press.

Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Goodrich, J., ... & Anders, J. (2023). Effective teacher professional development: new theory and a meta-analytic test. *Review of Educational Research*. doi: 00346543231217480.

Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research*, 78(3), 427-515.

Slavin, R. E., Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective reading programs for the elementary grades: A best-evidence synthesis. *Review of Educational*

- Research*, 79(4), 1391-1466.
- Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*, 79(2), 839-911.
- Stein, M. K., & Kim, G. (2011). The role of mathematics curriculum materials in large-scale urban reform: An analysis of demands and opportunities for teacher learning. In *Mathematics teachers at work* (pp. 57-75). Routledge.
- Stevenson, M., Hedberg, J. G., O'Sullivan, K. A., & Howe, C. (2016). Leading learning: The role of school leaders in supporting continuous professional development. *Professional Development in Education*, 42(5), 818-835.
- Stockard, J., Wood, T. W., Coughlin, C., & Rasplika Khoury, C. (2018). The effectiveness of direct instruction curricula: A meta-analysis of a half century of research. *Review of Educational Research*, 88(4), 479-507.
- Tallerico, M. (2005). *Supporting and sustaining teachers' professional development: A principal's guide*. Corwin Press.
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7), 3628-3651.
- Taylor, J. A., Roth, K., Wilson, C. D., Stuhlsatz, M. A., & Tipton, E. (2017). The effect of an analysis-of-practice, videocase-based, teacher professional development program on elementary students' science achievement. *Journal of Research on Educational Effectiveness*, 10(2), 241-271.
- Taylor, J. A., Kowalski, S. M., Polanin, J. R., Askinas, K., Stuhlsatz, M. A., Wilson, C. D., ... & Wilson, S. J. (2018). Investigating science education effect sizes: Implications for power

- analyses and programmatic decisions. *AERA Open*, 4(3), 2332858418791991.
- Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best Evidence Synthesis iteration (BES)*. New Zealand Ministry of Education.
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*.
- U.S. Department of Health and Human Services Office of Population Affairs (2023). *Core components of teen pregnancy prevention programs*.
https://rhntc.org/sites/default/files/resources/opa_tpp_program_components_brief_4-20-2023.pdf
- Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, 44(3), 519-558.
- Wanless, S. B., Patton, C. L., Rimm-Kaufman, S. E., & Deutsch, N. L. (2013). Setting-level influences on implementation of the Responsive Classroom approach. *Prevention Science*, 14(1), 40–51.
- Waters, J. T., & Marzano, R. J. (2006). *School district leadership that works: The effect of superintendent leadership on student achievement*. Mid-continent Research for Education and Learning (McREL).
- Weiss, M. J., Bloom, H. S., & Brock, T. (2014). A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, 33(3), 778-808.
- West, M. R., Morton, B. A., & Herlihy, C. M. (2016). *Achievement Network's Investing in*

- Innovation expansion: Impacts on educator practice and student achievement.* Harvard University: Center for Education Policy Research.
- Westbrook, A. L., Adams, C. T., & Taylor, J. A. (2023). Digging deeper into student–teacher–scientist partnerships for improving students’ achievement and attitudes about scientists. *The American Biology Teacher*, 85(7), 379-389.
- What Works Clearinghouse. (2021). Reporting guide for study authors: Group design studies. *Institute of Education Sciences*.
- Wilms, R., Mäthner, E., Winnen, L., & Lanwehr, R. (2021). Omitted variable bias: A threat to estimating causal relationships. *Methods in Psychology*, 5, 100075.
- Wilson, S. M. (2013). Professional development for science teachers. *Science*, 340(6130), 310-313.
- Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, 24(1), 173-209.

Table 1*Program Provider and Intervention Content for Students*

Category	Definition	Notes on how to report
Provider name, sector, and their relationship to the evaluation	Name the provider of the program and make clear what sector they work in (school staff, district staff, university, research firm, non-profit, for-profit). Describe relationship between providers and research/evaluation team, including whether the provider had input on research design or measures.	List.
Intervention content for students	Curricular subject, grade level, topic(s) covered (e.g., SEL routines), and pedagogy used to teach those topics (e.g., culturally responsive/sustaining teaching practices).	List all content addressed in the program as delivered to students, especially for integrated programs (e.g., focus on SEL and mathematics content, focus on student thinking). For programs with more than one content area, provide an estimate of relative emphasis.
Treatment contrast	To the extent possible, report on instructional improvement content experienced by control group students.	Same as above.

Table 2*Program Content for Teachers*

Category	Definition	Notes on how to report
Conceptual frameworks	Principles, standards, frameworks, or guidelines intended to inform teachers' instructional decision-making. Example: theories of teaching based in constructivist theories of learning.	Report relative emphasis of this vs. other types of program content by indicating whether it is a "major focus," "minor focus," "incidental focus," or "not a focus" as well as specifics regarding the opportunities to learn provided to teachers in this category. If category not mentioned, assumed to be unused in the program.
Bodies of knowledge	Knowledge intended to support teachers' instructional design and decision-making. Different from conceptual frameworks in that these are disciplinary or evidence-based bodies of knowledge. Examples: research on how students develop flexibility in solving math problems; science subject matter knowledge.	See above.
Prescriptions for practice	Guidance or materials that outline and support specific interactions between teachers and students in classrooms. Examples: lesson materials, including curricula (whether scripted or not); highly structured classroom routines.	See above.
Treatment contrast	To the extent possible, report on PD program content experienced by control group teachers.	Same as above.

Table 3*Program Professional Development*

Category	Definition	Notes on how to report
<i>Overall Duration</i>		
Maximum contact hours	Maximum number of PD contact hours available to teachers.	Sum of the total contact hours delivered across the PD formats described below.
Maximum time span	Maximum number of days/weeks/months teachers could have experienced the PD.	Total timespan over which PD activities were delivered (expressed in days, weeks, or months).
<i>PD Formats</i>		
Workshop(s)	Time for teachers to engage in PD outside of the regular school day, often off-site and often in significant blocks.	Report the maximum number of contact hours delivered. Specify whether it/they took place during the summer, after-school, or weekend—for example, teachers were offered a one-week summer workshop, comprised of 40 hours of content.
Coaching	Program provides or trains coach(es) to work with teachers. Coaching activities may include observation and feedback, reflection on practice, analyzing student performance, and representing and decomposing practice (e.g., via modeling).	Report the maximum number of contact hours delivered. For example, the program offered teachers opportunities for five 1-hour coaching sessions, one per week for 5 weeks. Specify the coaching activities included, e.g., observations of instruction with feedback, etc.
Professional learning communities	Time for teachers to engage in PD with colleagues, typically during a school day or briefly before/after a school day. Differs from workshops in that these tend to be more collaborative, site-based, and are often regularly scheduled on a weekly or monthly basis.	Report the maximum number of contact hours delivered. For example, the program facilitated twelve 1-hour grade-level team meetings.
Collective participation	Whether teacher attended PD with at least one other teacher from their school.	Report descriptively in a way that allows others to calculate the average concentration of teachers (e.g., 36 teachers attended from 18 schools).
Online	Whether PD was conducted in part or wholly online.	Report the maximum number of contact hours delivered. For example, the

program delivered 30 hours of asynchronous online PD. Report as synchronous or asynchronous.

Who/What Determines PD Content

Teachers	Defined as teacher-participants choosing or co-creating content with PD providers.	Report descriptively.
School leaders	Defined as school leaders (e.g., principal, instructional coach) selecting or co-creating the PD content.	Report descriptively. Where feasible, report the specific positions (e.g., superintendent, director of curriculum and instruction) of individuals that chose or co-created the curricula.
Analysis of student data	Occurs when an analysis of school/district student data shapes PD content.	Report descriptively.

PD Activities

Representations and/or decompositions of practice	Teachers study representations of teaching and/or break down teaching practices into constituent parts. Examples: viewing and analyzing a video of practice; reading and annotating case studies of strong practice.	Report relative emphasis of this vs. other activities in PD by indicating whether it is a "major focus," "minor focus," "incidental focus," or "not a focus" as well as specifics regarding the opportunities to learn provided to teachers in this category.
Approximations of practice	Teachers practice instruction outside of the classroom, often with peers playing the role of "student." Examples: teaching simulations; teaching demonstrations; rehearsals of practice.	See above.
Teacher reflection	Opportunities for teachers to reflect and gain insight on practice. Example: reflective journals, reflective coaching conversations.	See above.
Analysis of student work and/or performance	Analysis of student work or performance data. Examples: analyses of written student essays; videos of students solving problems; interim or state assessment data.	See above.
Planning instruction	Opportunities for teachers to integrate information or	See above.

materials from PD into their instructional practice.

Assistance in learning to implement curriculum materials	Teachers introduced to new materials, or study materials to strengthen implementation. Example: unit or lesson analyses.	See above.
--	--	------------

Teacher Participation

Mandatory or voluntary	Is participation mandatory, voluntary, or a mix of both for teachers?	Report as mandatory, voluntary, or a mix of both. Report whether teacher PD participation and attendance policies were set by the provider, school, district, or other entity.
------------------------	---	--

Attendance	Teacher attendance at PD activities.	Provide mean and range of teacher attendance, in hours. Report trends over time descriptively (e.g., attendance at PLCs waned during the year). If attendance varies between different program components (e.g., summer sessions and coaching sessions), report each separately.
------------	--------------------------------------	--

Teacher engagement	Teacher engagement with PD during sessions, including attentiveness, active participation in activities.	Report descriptively.
--------------------	--	-----------------------

Treatment contrast	Nature of PD experiences attended by control group teachers.	Report total hours and formats delivered to control group teachers. Report other elements above descriptively.
--------------------	--	--

NOTE: Please report any professional development planned in the original design but not delivered.

Table 4*Program Curriculum Materials*

Category	Definition	Notes on how to report
Curriculum contact hours	Percentage of a typical lesson the new curriculum materials were intended to replace, and percentage of the materials that were actually used.	Report the percentage of a typical lesson for which the new curriculum materials were intended to replace existing materials, and an estimate of the percentage of the intended curriculum materials that were actually used.
Curriculum time span	Time span over which the new curriculum materials were intended to be used, and time span over which the materials were actually used.	Report the length of time curriculum materials were intended to be used, and an estimate of the actual length of time they were used in the study (expressed in weeks or months).
Online content	Percentage of curriculum's content that is delivered to students online.	Estimate percentage of content that is intended to be delivered to students online. (If the actual implemented percentage differs, report this as well.)
Educative features	Materials contain guidance that would help teachers learn, for example, the content and/or disciplinary practices being taught, how students learn that content, or instructional practices to teach that content. Also includes design rationales meant to educate teachers about construction of curriculum.	Y/N with description.
Suggestions for adaptation	Materials contain explicit suggestions for lesson adaptation, for example, to the needs of learners or in response to time constraints.	Y/N with description.
Treatment contrast	Nature of the comparison group curriculum materials.	Report descriptively, and make sample lessons from the treatment and counterfactual curricula available in an online appendix where feasible.

Table 5*Program Support for Implementation*

To what degree did the program as implemented provide...	Definition	Notes on how to report
Student-facing materials	Student-facing materials that help teachers enact program-desired instruction. Includes prompts, tasks, curriculum materials, formative assessment items, lesson plans.	Y/N with description.
Implementation support meetings	Meetings to troubleshoot implementation—for example, allowing teachers to ask questions or share advice. Includes short meetings, phone calls, etc. led by program facilitators, their surrogates, or teachers themselves. These meetings do not share new program content.	Y/N with description.
Treatment contrast	To the extent possible, report on PD program content experienced by control group teachers.	Same as above.

NOTE: Please report any supports planned in the original design but not delivered.

Table 6

Setting-Level Resources, Contexts, and Barriers

Category	Definition	Notes on how to report
<i>District/School/Community Resources</i>		
Resources provided by the district	Kind and level of resources provided by the district. Examples include financial allotments for curriculum materials and program staff, additional time for professional development, meeting space.	Report descriptively and provide specifics of this support (if any). Report where feasible the specific positions of staff who provided the support (e.g., superintendent, vice superintendent for curriculum and instruction, school board, instructional services coordinator). Note also if any resources were needed but not provided.
Resources provided by the school	Kind and level of resources provided by the school. Examples include financial allotments for curriculum materials or program staff, school structures and routines that support readiness for the intervention; additional time for professional development, meeting space.	Report descriptively and provide specifics of this support (if any). Report where feasible the specific positions of staff who provided the support. Note also if any resources were needed but not provided.
Resources provided by the community	Kind and level of resources provided by the community. Examples include community expertise as well as community and family funds of knowledge that help shape program content and design; financial or in-kind contributions from local institutions.	Report descriptively and provide specifics of this support (if any). Note also if any resources were needed but not provided.
<i>Contexts for Implementation (May be Supports or Barriers)</i>		
School leader messaging	Degree to which school leader(s) either publicly endorsed the intervention, such as by messaging endorsement to teachers or participating in intervention activities themselves; or distributed negative or conflicting messaging about the intervention, such as by instructing teachers to focus on other initiatives or to only partially implement the program.	Report descriptively and provide specifics of this messaging (if any). Report where feasible the specific positions of those delivering the messaging (e.g., superintendent, vice superintendent for curriculum and instruction, school board, instructional services coordinator).

School climate	Degree to which the schools in which the intervention was implemented featured a supportive climate where teachers feel respected, experience collaborative relationships with colleagues, and have reasonable workload demands.	Report descriptively on facets of the school climate judged to be relevant to implementation (if any).
Teacher resources	Teachers' relevant prior experience, knowledge, attitudes, and/or other personal resources related to instructional change and the intervention content.	Report descriptively on facets of teachers' experience, knowledge, attitudes, and/or other personal resources judged to be relevant to implementation (if any).
<i>Barriers</i>		
Competing priorities or initiatives	Degree to which school/district has conflicting instructional guidance, competing classroom interventions, or competing priorities more broadly.	Report descriptively and provide specifics of these conflicts (if any).
Turnover of school and/or district leaders	Degree to which turnover of school and/or district leaders occurred during implementation, particularly turnover among program supporters.	Report level of relevant district and school leader turnover during program implementation, if any. Report where feasible the specific positions of those whose positions turned over.
Treatment contrast	To the extent possible, report any relevant setting-level differences between treatment and control groups that are predicted to affect the study outcomes.	Same as above.

Table 7

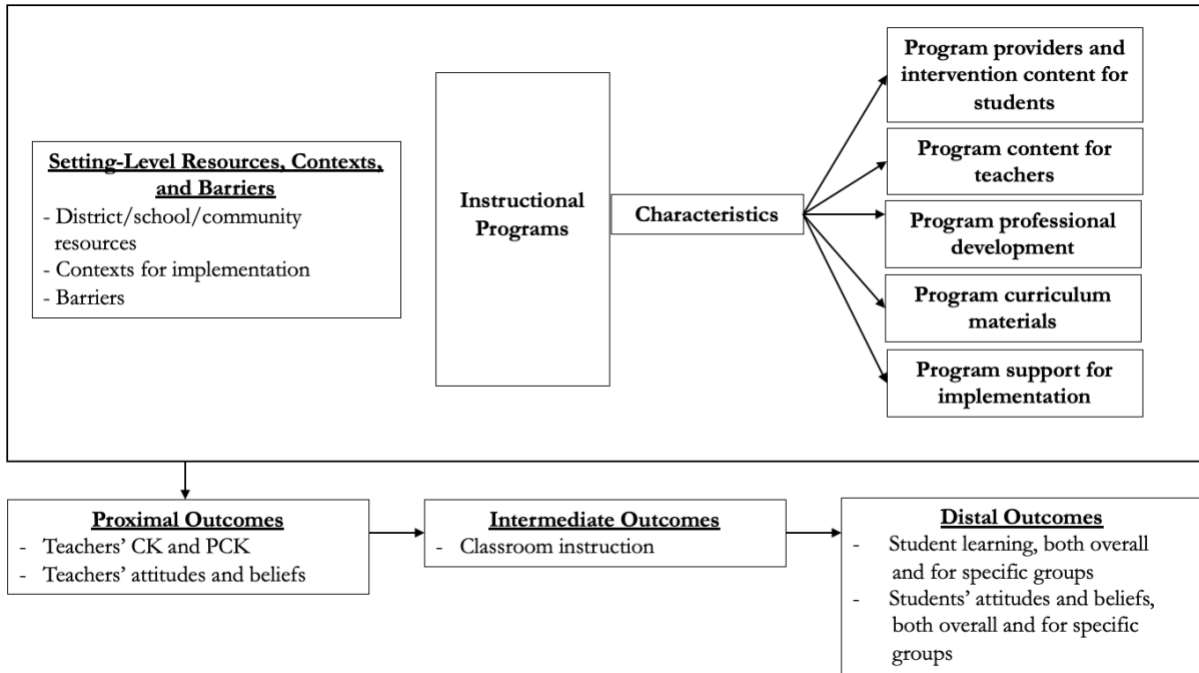
Example Summary Statistics Table with Overall and Specific Group Data for Outcome Measure

	Intervention N	Intervention Mean (SD)	Comparison N	Comparison Mean (SD)	ES (Var)
Overall	484	117.40 (51.85)	516	109.76 (51.58)	0.15 (0.004)
Students with IEPs	58	118.33 (56.63)	55	108.49 (46.98)	0.19 (0.036)
Students without IEPs	426	117.28 (51.23)	461	109.91 (52.14)	0.14 (0.005)

Note. N = sample size; SD = standard deviation; ES = standardized mean difference effect size (Cohen's *d* index); Var = variance of the effect size. Each ES represents the contrast for the row.

Figure 1

Logic Model for the Influence of Instructional Programs on Key Teacher and Student Outcomes (Adapted from Scher & O'Reilly, 2009)



Appendix A

Adapted from (redacted for review) to illustrate how authors might integrate the information requested in tables 2-5 into their publication. Text in bold indicates additional information added to the original report. Ellipses indicate non-relevant text from the original report omitted for the sake of brevity.

Sample and Setting

This study was conducted in one midsize school district serving a racially and socioeconomically diverse population of over 30,000 students across 46 school locations. Approximately 60% of the district's elementary schools are classified as Title 1. In addition, the area has experienced relatively high rates of student and teacher turnover due to a military presence in the surrounding community. This district has consistently performed below state averages in elementary mathematics.

The district's **superintendent approved the study and its** mathematics coordinator helped recruit 18 elementary schools to participate in the study, with a priority on recruiting schools that had (a) principals who were supportive of ambitious instruction and (b) varied student demographics. Eighteen principals agreed to participate in the study, allowing us to then recruit the fourth and fifth grade teachers slated to teach math in each of these schools during the first study year. Recruitment occurred in the summer and fall of 2010 and resulted in 88 teachers enrolling in the study (Cohort 1)... Table 1 shows descriptive statistics for these teachers. These teachers were 72% white and 9% male, and 55% held graduate degrees. They had, on average, almost nine years of teaching experience. During pre-service training, around half (57%) had taken more than two college-level math courses, 35% had taken more than two math content

courses for teachers, and 19% had taken more than two math methods courses. The baseline mathematical knowledge for teaching (MKT) scores of the teachers in the study were slightly below the national average for elementary teachers of all grades. **With the exception of the information about fidelity of implementation described below (see RQ1), no other setting-level resources or barriers were relevant to the implementation of the intervention.**

...

Study Design

After recruitment was finalized, teachers were randomized within their schools into treatment and control groups. Randomization resulted in 51 teachers (42 in Cohort 1 and 9 in Cohort 2) assigned to receive the professional development program (treatment group)...

As an incentive for participation, teachers in the control group were offered the opportunity to participate in district-designed science professional development in lieu of the mathematics professional development. The district received a lump sum financial contribution to hire an instructional coach to work with teachers on their science curriculum. The science professional development these teachers received was unrelated to the math professional development the teachers in the treatment group received and was unrelated to math instruction more generally, **focusing instead on inquiry-based methods for teaching science via workshops conveying knowledge about student science learning and representations of science teaching practice. All teachers also continued to participate in regular district professional development offerings beyond these two examples, and evidence from teacher surveys (seen in Table 5) suggest that such participation was balanced across groups.**

.....

Treatment

Program pedagogy and professional development. Teachers began their Math Thinking professional development with a four-day summer workshop in August 2010, with additional summer workshops held in each subsequent summer through August 2012. The institutes were held at a district facility and attendance was strongly encouraged. In addition to the summer institutes, Math Thinking staff administered between four (Year 3) and six (Years 1 and 2) one-day, in-person sessions during the course of the school year.^[1] In all, the program delivered over 40 contact hours per year over a three-year period, and involved participation of groups of teachers from the same grades, subjects, and schools. A separate training session was held each summer for school leaders and math coaches; these sessions featured mathematical problem-solving activities and observations of (and at times, participation in) the professional development sessions Math Thinking conducted with teachers. **Although an online component had been planned, as implemented all professional development was delivered entirely in person, and did not contain elements (e.g., coaching, implementation support meetings) cited in (Authors, under review) other than those noted explicitly above.**

Math Thinking staff determined the content of the professional development jointly with district staff, basing these decisions in part on student achievement data indicating areas of need. Conceptual frameworks were a major focus of the professional development; teachers were provided with a framework for “best practices” in mathematics instruction as well as a framework for analyzing student work, and discussed both frequently throughout the sessions. Improving teachers’ content and pedagogical content knowledge

was also a major focus; teachers typically started by working through a mathematical task or problem, often using multiple strategies or a concrete representation of the mathematical operation or idea. Then, one of several activities followed: a discussion of how students might solve the task in different ways; an examination of concrete examples of student work, for example, by watching a video of students solving the task, or interviewing actual students from nearby classrooms; a demonstration of a teacher teaching the task to her class; a discussion of how the task exemplifies best practices in classroom instruction; or a discussion anticipating issues that might arise as students encountered the task. **The last counts as a representation of practice, but the amount of video was small and teachers and facilitators only incidentally decomposed that practice, favoring a discussion about how the video exemplified program frameworks instead. Finally, teachers reflected on their own changes in practice and about best instructional practices more generally; this reflection, which took the form of journaling, constituted a minor focus of the program. No other pedagogies listed in (Authors, under review) appeared during the implementation of the program.**

An activity from a summer session illustrates how this occurred in practice. Teachers sitting in small groups were given a container of rice, tape, and three 5- by 8-inch index cards. The teachers were asked to make three shapes with the index cards—a circular cylinder, a triangular prism, and a square prism—by taping the two five-inch ends together with no gap or overlap. Once they had made the shapes, a Math Thinking staff member asked teachers to conjecture about which shape had the largest volume, or, alternatively, whether all shapes would hold the same amount. Many teachers argued that all the shapes would hold the same amount of rice

because they were the same height and the same distance around, having been made from the same-size index cards.

Teachers were then asked to test their hypotheses by actually filling the cylinder and prisms with rice and comparing the amount of rice used for each shape. Upon doing so, many were surprised that the circular cylinder actually held more than the other two shapes. In some groups, some teachers tried to recall and execute the formula for the volume of each shape, and in other groups, teachers tried to find the nets of the shapes; neither group made progress toward a reason for their outcome. The whole-group conversation that followed focused briefly on the mathematical reason for the cylinder's larger volume (the height was the same in all three shapes, but the areas of the bases differed) and then covered topics such as when to use the task during the school year, how long it should take to enact, and other desirable features of the task.

This activity sequence was typical of those we observed during both the summer and within-year sessions. The activities combined teachers collaboratively solving mathematics problems and discussing best practices in classrooms, ranging from mathematical practices (e.g., conjectures), to new forms of student assessment (e.g., interviews), to new kinds of classroom activities and pedagogical techniques. In some sessions, fourth and fifth grade teachers were separated into different rooms to work on grade-level specific content, or to make connections between district instructional guidance and Math Thinking materials.

Program-provided student-facing materials. Both treatment and control teachers used the same district-provided curriculum, **Harcourt Mathematics**. In addition, treatment group teachers were given books, often published by Math Thinking, with lesson ideas or detailed explanations of the mathematics that teachers might teach. Many of the activities that the Math

Thinking staff used in training came from the books the teachers received, and teachers were given opportunities to connect Math Thinking activities and materials to district pacing guides and the state test. **Together, these books contained over 100 complete mathematics lessons, but were not meant to form a coherent curriculum; instead, program developers asked teachers to augment their existing curriculum with these lessons once or twice a week throughout the year. These materials contained no educative features for teachers, no design rationales, and no suggestions for adaptations based on student need. The program contained no online component for students.**

...

Results

RQ1: Fidelity of Implementation

As shown in Table 5, fidelity of implementation was strong during the first year of the study, but declined in the subsequent two years. During the first year of the study, Math Thinking staff spent considerable time engaging with the district, **determining program content** in a way that would meet the needs of the district. However, after the district's math specialist retired at the end of the first year of the study, it was difficult **for program staff** to reengage with district leadership in the same way. The district was in a state of flux, with new state and district standards, new assessments and a new superintendent and as a result, the district staff appeared pulled in many different directions.

Professional development sessions in all three years were led by Math Thinking trained team members and, as detailed in Table 5, observations conducted by study team members suggest that the professional development was consistent with the description of the program

goals and pedagogical methods in all three years. Attendance at the training was high during the first year of the program, when 35 of the 42 teachers (83%) randomly assigned to the treatment group attended the summer training and 90% (38 of 42) of the teachers participated in all six of the in-service training sessions that were held. However, attendance at the professional development dropped in the subsequent two years. During the second year, on average, only 69% of the teachers attended trainings; in the third year, only 55% of the teachers remaining in the treatment group attended. **Teacher engagement during workshops was high in year 1, and continued to be high among attending teachers in the subsequent two years.** The week-long summer institute was shortened to 3 days in Year 2 and only 4 in-service days were delivered in the last year of the study.

In the second and third study years, principal participation and engagement with the professional development also declined. During the first year of the study, most principals attended the special session of the professional development designed to help them understand the training their teachers were receiving. During the third year, only two attended.

[1] Distance learning and coaching were proposed by the Math Thinking team but not taken up by the teachers who were participating in the study.