Effective Implementation Cohort

Cohort Wide Learning Plan

August 2022

Prepared for the Bill & Melinda Gates Foundation by

Caryn S. Ward, Ximena Franco-Jenkins

NIRN | NATIONAL IMPLEMENTATION RESEARCH NETWORK
FRANK PORTER GRAHAM CHILD DEVELOPMENT INSTITUTE

CONTENT

EXECUTIVE SUMMARY

PURPOSE

The overall aim of the Effective Implementation Cohort (EIC) investment is to increase district capacity to implement a high-quality middle years math curriculum as part of a Coherent Instructional System (CIS) to accelerate learning for students who are experiencing poverty, Black, Latino/a, and/or English Learner (EL)-Designated ("priority students"). As a Learning Partner, NIRN seeks to support partnerships between Providers and Local Education Agencies (LEAs) in their implementation and measurement efforts, as well as to collect and study data to answer the investment's learning questions within the cohort wide learning agenda. Implementation support activities and data collection, analysis, and reporting activities for the cohort wide learning agenda will be guided by the Cohort's learning questions and theory of action grounded in implementation science.

THE BILL AND MELINDA GATES FOUNDATION'S LEARNING QUESTIONS ARE GROUPED INTO FOUR CATEGORIES:

APPLICATION OF ESTABLISHED PRACTICE

ENABLING CONTEXTS

VARIABILITY IN IMPLEMENTATION

MEASUREMENT

DESIGN

Our research design examines how degrees of and factors related to implementation influence a teacher's ability and beliefs to implement a curriculum as intended (fidelity), and ultimately, how teacher self-efficacy and the level of fidelity of implementation affect students' mathematics engagement, experience, beliefs, and achievement. A mixed-methods approach will be used to measure the strength and variation of implementation supports across Provider-LEA partnerships (n = 19) at the district, school, and classroom/teacher levels in order to identify cohort-wide trends and establish relationships between implementation outcomes and other contextual factors (i.e., locale, governance model, and leadership stability).

Various implementation supports are provided by NIRN, Providers, and LEAs. Multiple methods of data collection will be used to gather information from LEA and school leaders, teams (inclusive of district and school staff), math teachers, instructional coaches, students, and providers. Specifically, data collection methods will include teacher and student surveys, observations, interviews, administrative data, and product reviews. Common data collection methods across the cohort will be used to examine students' beliefs, experience, and engagement in mathematics (i.e., student survey) and the LEA's and school's organizational capacity (e.g., District Capacity Assessment, Principal Implementation Leadership Survey). For other implementation outcomes such as teacher knowledge, self-efficacy, and practice, Providers and LEAs will be using instruments that best fit their individual work and context.

Thus, the analysis will be largely descriptive using cohort-wide trends to achieve goals including:

Creating evidenced-based models for the implementation of high-quality middle years math curricula

Adding to the knowledge base available to education professions on how to implement high-quality middle years math curricula, particularly for students of color and those experiencing poverty

Providing participating Providers and LEAs with meaningful, formative feedback to inform their practice

DEEP DIVE IMPLEMENTATION STUDY

To further investigate the learning question of how does implementation affect student math learning and for whom and in what contexts, student and teacher level unit of analysis will be conducted with a select number of Provider-LEA partnerships (i.e., up to 4 to 6 LEAs). The sub-sample may include LEAs also participating in the AMS study of efficacy and enactment, depending on the final district/school selection and the student and teacher data being collected as part of that study, as well as those LEAs with the ability to provide linked student and teacher administrative data. Efforts will be taken to ensure the sub-sample of LEAs is representative of key contextual factors such as representation of priority students and LEA locale. Quantitative analyses for this sub-sample will utilize propensity score analysis and multilevel linear modelling. Although limitations to generalizability are present with this deep dive approach, controlling for school and teacher factors that might influence implementation fidelity and student outcomes will allow the isolation of the effect of implementation on student learning, for whom, and in what contexts. Given the purpose of this investment is an implementation learning agenda and does not seek to establish causal relationships, a control group will not be used.

OUTPUTS

In addition to improving LEA and Provider capacity, the EIC has the goal of producing reliable, practical evidence and measures to inform planning and implementing a district-wide, high-impact math improvement initiative. The evidence and measures will support:

Providers to design more effective implementation services

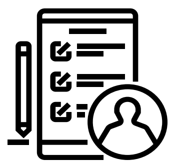
Districts as they track implementation progress

The field to better understand how high-quality implementation affects enactment, and, in turn, student learning

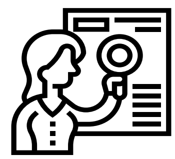
A number of potential outputs will be generated from the EIC for use by the field of education and in other BMGF investments related to the implementation of middle years mathematics curriculum as part of a Coherent Instructional System.

Specifically, the design will use the school level as the primary unit of analysis to examine all of the learning questions. Qualitative and quantitative data will be conducted through descriptive, visual trends, and thematic analyses. **Example outputs of the cohort wide learning agenda using this design include:**

Definition of readiness conditions at school and district levels that support scaling effective solutions



A curated list of self-assessment tools for district leaders to gauge their system's readiness and identify next steps to prepare for implementation



Identification of context-specific readiness factors that leaders in priority contexts should consider

INTRODUCTION

The overall aim of the EIC investment is to increase district capacity to implement high-quality middle years math curricula as part of a Coherent Instructional System to accelerate learning for students who are experiencing poverty, Black, Latino/a, and/or English Learner (EL)-Designated ("priority students"). As a Learning Partner, NIRN seeks to support partnerships between Providers and Local Education Agencies (LEAs) in their implementation and measurement efforts, as well as to collect and study data to answer the investment's learning questions within the cohort wide learning agenda.

THE BILL AND MELINDA GATES FOUNDATION'S LEARNING QUESTIONS ARE GROUPED INTO FOUR CATEGORIES:

- **APPLICATION OF ESTABLISHED PRACTICE**
- **ENABLING CONTEXTS**
- **VARIABILITY IN IMPLEMENTATION**
- **MEASUREMENT**

It is important to note that the EIC investment consists of two phases.

Phase I (January-July 2021)
Was focused on implementation planning and creating readiness for high-quality middle years mathematics curriculum.

Phase II (October 2021- December 2024)
Is focused on implementation of the high-quality middle years math curriculum.

RESEARCH DESIGN

Learning Questions

This investment aims to answer questions on the application of established practice and enabling contexts beginning in Phase I and continuing to answer those questions in Phase II. The learning questions related to variability in implementation and measurement will be answered during Phase II.

The specific questions under each category are detailed below:

APPLICATION OF ESTABLISHED PRACTICE

1. Which lessons learned from implementation research (e.g., curriculum enactment studies, education and health implementation studies) scale across this set of partnerships? Which do not?
2. What do educators and system actors identify as the reasons why evidence-based ideas scale or do not scale?

ENABLING CONTEXTS

1. What conditions most distinguish districts' and schools' readiness for, approach to, and success with implementation?
2. What conditions maximize likelihood of successful implementation across contexts relevant to priority students?
3. Which conditions are go/no-go factors versus important readiness conditions that a provider can help establish?
4. How does the set of key enabling conditions vary across priority contexts?
5. How have districts' priorities for implementation of new curricula or instructional systems shifted in light of COVID-19? How do these shifts influence key enabling conditions?

VARIABILITY IN IMPLEMENTATION

1. What is the “menu” of interventions taken by systems to support implementation? Are some seen as more promising and likely to be taken up than others? Which seem to achieve successful implementation most often? How does this vary by context?
2. What are the challenges for implementation faced across grantees? What patterns are there to those challenges, especially as related to contextual factors (e.g., student composition, characteristics of previous instructional system, teacher characteristics, broader enabling conditions)?
3. What best practices emerge across the dimensions of implementation support that lead partners are being asked to provide, as well as any others that arise? How are those best practices related to patterns in contextual factors as described above?
4. How do the findings from this cohort reinforce, complement, or differ from the findings about classroom-level factors that drive curricula’s effectiveness in the separate enactment studies that the Foundation is funding?

MEASUREMENT

1. What defines effective implementation? What are evidence-based leading indicators, lagging indicators, and data sources to assess progress and impact?
2. What are the best-established measures for assessing readiness to take on implementation, self-assessment of what is needed to support implementation, and factors (e.g., system conditions) that should be tracked to support implementation?
3. How do grantees use or eschew established leading and lagging measures related to both CIS and implementation to identify challenges and make improvements to systems of support? Why? We are particularly interested in findings related to the ability of systems to collect and share data, perceptions of the validity of measures, and actionability of data.
4. What factors not identified in established measures are critical to identifying challenges and making improvements? How do districts and school systems identify these factors?
5. How, and how much, did implementation affect student math learning and/or leading indicators of math learning? In what contexts and for whom?
6. How does teacher perception (e.g., feasibility, acceptability, and appropriateness) of the math curriculum within CIS influence implementation?

RESEARCH DESIGN

Implementation Supports

Our research design is descriptive and observational in nature. The EIC seeks to answer a set of cohort-wide learning questions with regards to the relationship between enabling context, professional learning and high-quality mathematics curricula, teacher outcomes, and student outcomes. Both qualitative and quantitative data will be used to answer the cohort-wide learning questions. This mixed-methods approach will be used to measure the strength and variation of implementation supports across Provider-LEA partnerships (n = 19) at the district, school, and classroom/teacher levels in order to identify cohort-wide trends and establish relationships between implementation outcomes and other contextual factors (i.e., locale, governance model, and leadership stability).

The EIC will explore and examine aggregate trends and statistical relationships between district readiness and capacity, teacher practice and self- efficacy, and student achievement, engagement, experience, and beliefs in mathematics. Specifically, the goal of the EIC is to create practical evidence and measures to inform planning and implementing a district-wide, high-impact math improvement initiative in service to supporting middle school Black, Latino/a, English Language Learners, and students experiencing poverty.

The three levels of implementation supports are:

NIRN'S SUPPORT TO PROVIDERS AND LEAS AS THE LEARNING PARTNER **1**

PROVIDERS SUPPORT FOR LOCAL EDUCATION AGENCIES **2**

PROVIDER/LEA SUPPORT FOR TEACHERS AND SCHOOL LEADERS **3**

RESEARCH DESIGN

Theory of Action

A cohort wide theory of action will be used to examine the cohort's learning questions. Specifically, the theory of action outlines how change will lead to improved practices and outcomes and guides the development of hypothetical statements. The proposed theory of action is based on implementation science that posits three factors are necessary for improved outcomes including: 1) effective practices (e.g., high-quality middle years math curriculum and instructional practices), 2) effective implementation (e.g., high quality professional learning services), and 3) enabling context (e.g., linked implementation teams and use of improvement cycles) (Fixsen et al., 2013). See **Figure 1** for the Theory of Action.

FIGURE 1.
Theory of Action

IF	THEN	THEN	THEN
Establish linked teams to create **enabling conditions using sciences** of implementation and improvement at the Provider, district and school levels	Districts and schools will have increased **capacity** to select and implement High-Quality Middle years Math Curriculum & Professional Learning Supports	Teachers will have improved **self-efficacy** and **integrity** of implementation for the High-Quality Middle years Math Curriculum	Students who are Black/ Latino/a and/ or affected by poverty will have improved **math achievement, enjoyment, self-efficacy, and growth mindset**

Using the theory of action, a model outlining the enabling conditions (i.e., readiness indicators) and series of short-term, intermediate, and long-term outcomes has been developed to inform the learning questions (see Figure 2 p.14). In addition to the readiness indicators, a number of district context factors have been identified to be examined.

FIGURE 2.
Map of Theory of Action & Outcomes

Phase	Learning Questions	Component	Term	Outcomes		
PHASE 1	LQ1 Application of Established Practice; LQ2 Enabling context	**IF** Enabling Conditions	Readiness Indicators	District Teaming LEAs Executive Sponsor's Engagement Communication Assessing Fit and Feasibility	Implementation Planning Measurement Planning	
PHASE 2	LQ1 Application of Established Practice; LQ2 Enabling context; LQ3 Variability in Implementation; LQ4 Measurement	**THEN** Professional Learning Services (PLS) & High-Quality Math Curriculum	Short Term	Acceptability Adoption Appropriateness Feasibility Quality of PLS	Organization Capacity: Communication Data System Leadership Quality Resource Allocation Stakeholder Engagement Supporting Policies & Procedures	
			Intermediate	Cost Integrity/Fidelity Penetration/Reach Sustainability		
		THEN Teacher Outcomes	Intermediate	Teacher Practice: Fidelity/ Integrity	Teacher Self-Efficacy: Math Curriculum Cultural Pedagogy	
		THEN Student Outcomes	Long Term	Student Math Achievement (benchmark, formative and summative assessments)	Student Beliefs (enjoyment, self-efficacy, and growth mindset)	Student Engagement (academic, social & emotional)

District Contextual Factors:

Locale

Size

Demographics of Student Population

Governance Structure (i.e., elected vs appointed, relation with school board)

LEA Leadership Stability (i.e., Superintendent, Chief Academic Officer)

Decision Making Model (i.e., centralized or decentralized - site based)

School Leadership Stability

Teacher Retention

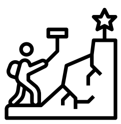
Labor Relations (leadership & union agreement on strategy)

Community Relations (engagement of school board, engagement of regional/state support)

Financial Management i.e., funding model - braided, transferred, blended)

SUMMARY OF EMPIRICAL EVIDENCE

A summary of the existing empirical evidence for different theory of action constructs is provided. In addition to highlighting the evidence of what is known, we also highlight how the proposed study will contribute to or generalize new knowledge regarding these constructs for the field.



ENABLING CONDITIONS

IF

If linked teams are established, they will create enabling conditions to support the use of implementation improvement at the Provider, district, and school levels.

Our theory of action begins with the premise that the foundation for improved outcomes for all students in mathematics is the establishment of linked teams and the creation of enabling conditions at district and school levels to support the use of a high quality middle years mathematics curriculum within a coherent instructional system. Using the sciences of implementation and improvement, we will examine the role and functioning of a linked teaming structure (provider - district - school - level teaming) to create and maintain the necessary readiness conditions for implementation.

Readiness is defined as a developmental point at which a person, organization, or system has the **capacity** and **willingness (momentum)** to engage in a particular activity (Fixsen et al., 2013; Wandersman & Scaccia, 2018).

"Capacity" *is defined as the perceived abilities, skills, and expertise of school leaders, teachers, faculties to execute or accomplish something specific, such as leading a school-improvement effort.*

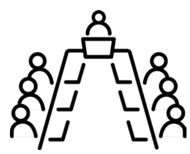
"Willingness" *is defined as the quality or state of being prepared to do something.*

Readiness is a condition that needs to be developed and maintained throughout implementation. It is often considered to be a precursor to successful implementation but has not been as widely researched as individual readiness for change (Weiner, 2020).

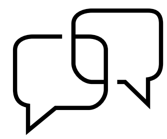
Social cognitive theory and motivational theory supports the following hypotheses underlying our theory of action: **When organizational readiness for change is high, organizational members are more likely to initiate change, exert greater effort to implement the change, and persist in change implementation despite obstacles** (Gist & Mitchell, 1992; Meyer & Herscovitch, 2001).

Specifically, Provider-LEA pre-implementation readiness conditions will be examined across six constructs:

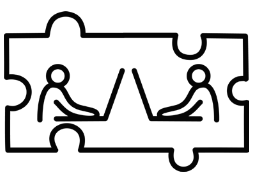
District Teaming (DT)



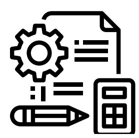
LEA Executive Sponsor Engagement (ES)



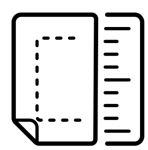
Communication (C)



Assessing Fit and Feasibility (FF)



Implementational Planning (IP)



Measurement Planning (MP)

Each readiness construct consists of 4 to 6 operationalized indicators contextualized for this specific investment (See **Appendix B** for all readiness construct definitions and indicators). A summary of the evidence base for each of these readiness constructs is provided in **Table 1 (p.18)**.

TABLE 1.

Readiness Constructs' Definitions & Rationales

Throughout Table 1, the term "**critical perspectives**" is used to refer to individuals and groups including, but not limited to, school staff, students, families, and community partners.

Readiness Construct

District Teaming

Definitions	A representative team has been formed at the district level and is working to lead implementation and create the enabling conditions in selected schools so that educators can make full use of high-quality middle years math curriculum.
Rationale	Research has shown that using implementation teams to actively and intentionally make changes produces higher rates of success more quickly than traditional methods of implementation with less active approaches (Higgins et al., 2012; Metz et al., 2015). Research has also shown implementation teams have a significant impact on whether evidence based practices are implemented and sustained over time (Leithwood & Azah, 2017; McIntosh et al., 2018). Furthermore, research shows the importance of a shared resolve among an implementation team, noting that implementation as a "team sport" requires each member to demonstrate and share commitment to the work (Nilsen, P., & Birken, S. A. 2020). For mathematical instructional systems specifically, evidence exists to support the positive impact of school based professional learning communities and teacher networks on teacher learning (Cobb & Jackson, 2011).

Readiness Construct

LEA Executive Sponsor Engagement

Definitions	The Executive Sponsor(s) champions and supports district and school staff as they engage in implementation of high-quality middle years mathematics curriculum. An executive sponsor is an individual with the ability to influence others, authority to make decisions regarding resource allocation, institutional knowledge, time, and positive relationships with staff and critical perspectives.
Rationale	Executive sponsors or champions are defined as those individuals who are internal to the organization, have an intrinsic interest in the needed change, are driven to succeed, have a positive approach to the work, and a strong conviction or belief in their ability to do the work. Research has found that champions represent a "necessary but not sufficient" condition for implementation success. Champions alone are inadequate to bring about change, yet in combination with other factors were found essential to implementation success (Miech et al., 2018). The role of a "champion" or executive sponsor at the LEA level has not been researched specifically for mathematics. Instructional leadership provided by principals has been found in research to be critical for successful implementation (Cobb & Jackson, 2011; Katterfeld, 2013).

Readiness Construct
Communication

Definitions	Frequent and accurate information regarding implementation planning and progress is exchanged between critical perspectives and acted upon by the identified persons. Goals of communication include sharing information, gathering feedback and input, clarifying expectations, and celebrating successes.
Rationale	Research has demonstrated that strategic communication is an important driver for system change leading to improved performance (Fullan, 2010; Levine, 2014; Duffy & Chance, 2006).

Readiness Construct
Assessing Fit and Feasibility

Definitions	An assessment conducted by districts to better understand how a new or existing high-quality middle years mathematics curriculum works within their existing context to support implementation planning and use.
Rationale	Based on social cognitive theory, change efficacy is largely a function of organizational members' cognitive appraisal of task demands, resource availability, and situational factors (Gist & Mitchell, 1992). "When organizational members share a common, favourable assessment of task demands, resource availability, and situational factors, they share a sense of confidence that collectively they can implement organizational change" (Weiner, 2020, p. 222). Although the importance of contextual variables and fit is often referenced in selection and implementation of evidence based practices, there is lack of consensus on specific elements that constitute contextual fit and lack of a strong research base (Horner, Blitz, & Ross, 2014). Research has shown that districts consider most prominently the three factors of alignment to standards, impact on student achievement, and cost effectiveness when selecting curriculum materials (Allen & Seaman, 2017; Campbell & Polikoff, 2017; Zeringue et al., 2010).

Readiness Construct
Implementation Planning

Definitions	Comprehensive implementation strategies are specified within a plan to ensure capacity is developed to support successful use and sustainability of a high-quality middle years mathematics curriculum.
Rationale	Powell and colleagues (2015) found having a formal implementation "blueprint" to guide implementation was a key implementation strategy for change. The plan or blueprint outlines the 1) aim/purpose of the implementation, 2) scope of the change, 3) timeline for change, and 4) appropriate performance/progress measures. Within K-12 education, quality school planning has been found to be associated with positive implementation outcomes (Strunk et al., 2016).

Readiness Construct
Measurement Planning

Definitions	Utilize multiple methods to collect and review data to: (1) inform decision-making for continuous improvement, (2) examine effectiveness, and (3) communicate with critical perspectives-.
Rationale	Measurement planning and data usage have been found to be a district practice associated with successful implementation of evidence based practices (Byrk et al., 2015; Fullan & Quinn, 2016; George et al., 2018; Leithwood & Azah, 2017).

In addition to the identified readiness constructs, a number of district and school demographics and contextual factors (such as locale, governance structure, labor relations, community relations, and financial management) will be examined. Student and teacher demographic and socio-economic status has been found to be associated with student achievement (LaCour & Tissington, 2011; McCoy, 2005).

Furthermore, as the percentage of students living below poverty and the diversity among students increases, teachers' perceptions of shared norms for instruction, climate, openness with parents, and district support decreases. Although larger urban districts have been found to foster greater shared leadership among teachers and staff, perceptions of school climate, openness with parents, and district support decreases as compared to teachers' perceptions within suburban and rural districts (Louis et al., 2010).

In terms of students having a teacher of the same race, Redding (2019) found that students' perceptions of a same-race teacher is more favorable. This relationship varies by the school level. There was also strong evidence that Black students perform higher on academic achievement tests when assigned to a Black teacher. There was less evidence of this relationship for Latino/a students.

IF

THEN

PROFESSIONAL LEARNING AND MATH CURRICULUM IMPLEMENTATION OUTCOMES

IF/THEN Districts and schools will have increased capacity to select and implement high-quality middle years mathematics curriculum and professional learning supports. While significant financial and human resources are allocated to professional learning activities in districts, less is allocated to support the foundational components that facilitate implementation and scale-up. As displayed in **Table 2**, Proctor and colleagues (2011) identified a set of outcomes to understand what and how processes work to facilitate or create barriers to implementation. While conceptually sound, they have not undergone rigorous examination in education. The EIC body of work adds significantly to how districts and schools can better attend to, collect, analyze, and use implementation outcomes to increase the likelihood that teachers implement practices as intended.

TABLE 2.
Implementation Outcome Constructs' Definitions & Rationales

Implementation Outcome Construct	**Acceptability and Appropriateness**
Definitions	Acceptability is defined as the perception that a given practice (i.e., math curriculum) or service (i.e., professional learning services) is agreeable and satisfactory.
	Appropriateness is defined as the perceived fit, relevance or compatibility of a practice (i.e., math curriculum) or service (i.e., professional learning services) to address a specific need.
Rationale	While little research has established their predictive capability as implementation outcomes, Weiner and colleagues (2017) developed measures of the acceptability and appropriateness of interventions and practices in local contexts. Their work identified acceptability of an intervention or practice as being approved, appealing to, liked, and welcomed by practitioners. Furthermore, appropriateness was operationalized as being fitting, suitable, applicable, and a good match to the practitioners current work and environment.
Implementation Outcome Construct	**Feasibility**
Definitions	Feasibility is the extent to which a practice (i.e., math curriculum) or service (i.e., professional learning services) can be successfully delivered or used within a given context.
Rationale	Lyon and others (2019) found that beliefs on the importance and feasibility of implementation of a practice can shift over time and should be addressed in implementation planning.

Implementation Outcome Construct

Quality of Professional Learning

Definitions	The extent to which a service is perceived to be of high quality and usefulness.
Rationale	Quality of professional learning had the potential to impact both teacher practice and student outcomes. Brock and Carter (2016) found that modeling and performance feedback during training had a significant positive impact on fidelity, especially when it came to implementing programs for students with disabilities. Additional evidence has shown that job embedded training increases teacher-self-efficacy and student math achievement (Althauser, 2015).

Implementation Outcome Construct

Organizational Capacity: Communication, Data Systems, Leadership Quality, Resource Allocation, Stakeholder Engagement, Supporting Policies & Procedures

Definitions	The systems, activities, and resources that are necessary for schools to successfully adopt and sustain practices (i.e., math curriculum) such as Communication, Data System, Leadership Quality, Resource Allocation, Stakeholder Engagement, and Supporting Policies & Procedures.
Rationale	Organizational capacity is critical to understanding the resources, processes, and structures that support teachers in implementing new practices and programs. Organizational activities such as communication, leadership, stakeholder engagement, and supportive policies and procedures that are aligned to the literature that suggests schools can enhance teacher's involvement in professional learning and connection to the organization (Sleeger et al., 2014). Additionally, Malen and colleagues (2015) suggest "building capacity requires the timely allocation of appropriate levels and types of foundational resources and the strategic construction of the organizational conditions that allow those resources to be put to productive use."

Implementation Outcome Construct

Cost/Sustainability

Definitions	Cost of implementation efforts for the practice (i.e., math curriculum) or service (i.e., professional learning services)
	Sustainability is the extent to which a practice (i.e., math curriculum) or service (i.e., professional learning services) is maintained within a system
Rationale	Cost as an implementation outcome is important to both sustainability and replicability. Levin and Belfield (2015) suggest that evaluations should include a component that collects and analyzes cost information using the ingredients method, which specifies the activities that are needed to fully enact the program or practice. Estimate cost figures allow researcher to establish cost-effectiveness and cost-benefit estimates that can inform decision making.

TEACHER OUTCOMES

THEN

THEN Teachers will have improved self-efficacy and integrity/fidelity of implementation of high-quality middle years mathematics curriculum.

The EIC theory of action posits with improved organizational capacity and enabling conditions in place, teachers will experience positive changes in their knowledge and skills in the delivery of mathematics instruction using the high quality curriculum materials as well as in their confidence in delivering instruction and meeting student needs including the use of culturally pedagogy strategies. Overall, research has shown that implementation strategies such as professional development and coaching can have an impact on teacher beliefs and practice (Desimone & Garet, 2015; Kraft et al., 2017). Furthermore, teachers' beliefs about instruction and student learning influences their instructional practice as well as their response to change initiatives (Spillane, Hopkins & Sweet, 2017; Stipek, Givvin, Salmon, & MacGyvers, 2001). Teacher outcomes are displayed in **Table 3**.

TABLE 3.
Teacher Outcome Constructs' Definitions & Rationales

Teacher Outcome Construct
Teacher Knowledge & Skills of Math Instruction

Definition	The degree of knowledge and skills in teaching mathematics (delivery of instruction).
Rationale	Campbell and colleagues (2014) found that mathematical content and pedagogical knowledge of middle-grades teachers were each directly and positively related to their students' mathematics achievement.

Teacher Outcome Construct

Fidelity/Integrity

Definition	The extent to which a practice (i.e., math curriculum) is delivered as originally developed and specific in plans and protocols.
Rationale	Hill and Erickson (2020) found in their review of evidence for fidelity of implementation of STEM curriculums and classroom interventions that better fidelity correlated with better program outcomes. In addition, Hill and Erickson found that classroom observations tended to see more positive fidelity outcomes than teacher self-report. Similarly, Ysseldyke and colleagues (2003) found that math students in classes of high-implementers demonstrated greater academic performance than students in the control group.

Teacher Outcome Construct

Self-Efficacy: Math Instruction

Definition	The degree of confidence in delivering math instruction and meeting student needs
Rationale	Recent research, however, has documented a positive relationship not only between direct measures of teachers' knowledge of mathematics content and pedagogy and student achievement (Baumert et al., 2010; Hill, Rowan, & Ball, 2005) but also between teachers' beliefs about mathematics teaching and learning and student achievement (Love & Kruger, 2005; Clark, 2014). Teacher self-efficacy has been associated with positive outcomes for both teachers and students. In terms of teacher outcomes, it is a significant predictor of a teacher's understanding of students (Caprara, Barbaranelli, Steca, & Malone, 2006), is associated with a teacher's ability to more effectively respond to students' needs (Hoy & Spero, 2005) and is associated with the development of meaningful curriculum and learning opportunities in the classroom (Reyes, Brackett, Rivers, White, & Salovey, 2012).

Teacher Outcome Construct

Teacher Self-Efficacy: Cultural Pedagogy

Definition	The degree of confidence in using culturally pedagogy strategies
Rationale	Studies indicate that students make academic progress when their teachers are prepared to incorporate culturally relevant teaching strategies into classroom pedagogy (Cammarota & Romero, 2011), and that when teachers recognize and acknowledge students' linguistic and cultural backgrounds and incorporate them into learning, students experience greater academic success (Bui & Fagan, 2013; Dee & Penner, 2017).

THEN

STUDENT OUTCOMES

THEN Students who are Black, Latino/a, and/or affected by poverty will have improved math student achievement, enjoyment, self-efficacy, and growth mindset.

The EIC Theory of Action acknowledges the complex nature of student-centered learning. Many system, community, school, teacher, family, and student factors influence what students believe about their own ability to do mathematics, how they engage in mathematics coursework, and ultimately, how they perform on local and state benchmark assessments.

TABLE 4.
Student Outcome Constructs' Definitions & Rationales

Student Outcome Construct **Beliefs & Self-Efficacy**	
Definition	For beliefs and self-efficacy, students' identity, confidence, and understanding that productive struggle and mistakes are opportunities for learning and growing their mind will be examined. In addition, students' experience of safety, respect, belonging, inclusion, and joy in the classroom/math learning will be examined.
Rationale	Student mathematics self-efficacy and growth mindset has been shown to impact student mathematics performance and enrollment in more rigorous mathematics classes (Evans, 2015; Warwick, 2008; Yeager et al., 2019). In addition, the level of mathematics directly impacts math anxiety and growth mindset (Huang et al., 2019). Long-term, STEM identity mediates the effect of STEM mindset and career interest (Cribbs et al., 2021).

Student Outcome Construct **Engagement**	
Definition	Engagement is defined as the level of motivation and participation that students display while learning mathematics.
Rationale	Defined through cognitive, academic, behavioral, and affective engagement, significant evidence exists linking student engagement to academic outcomes (Finn & Zimmer, 2012; Fredricks & McColskey, 2012). Furthermore, existing achievement gaps between children of color and white children are often thought of being a product of difference in engagement - a conclusion not supported by the literature (Bingham & Okagaki, 2012). Skilling and others (2015) highlight the need for attending to student engagement for students in middle school mathematics classes. Less is known about how implementation fidelity/integrity and teacher self-efficacy influence student mathematics engagement.

Student Outcome Construct **Math Achievement**	
Definition	Math achievement is defined as the extent to which students develop a deep knowledge of math and achieve educational goals/standards.
Rationale	Math achievement can be influenced by a number of different factors. For example, Crawford (2012) found that fidelity of structure (intervention specific fidelity) had a positive effect on student mathematics achievement for middle school students. Additionally, students attending schools with high levels of professional learning activities exhibited more growth than those attending school with lower levels of professional learning activities (Allensworth et al., 2021).

RESEARCH DESIGN

Outputs

In addition to improving LEA and Provider capacity, the EIC has the goal of producing reliable, practical evidence and measures to inform planning and implementing a district-wide, high-impact math improvement initiative.

The evidence and measures will support:

Providers to design more effective implementation services

Districts as they track implementation progress

The field to better understand how high-quality implementation affects enactment, and, in turn, student learning

A number of public goods (i.e., outputs) will be generated from the EIC for use by the field of education and in other BMGF investments related to the implementation of a high-quality middle years mathematics curriculum within a Coherent Instructional System. See table and figure below for a timeline of the outputs to be generated.

AT A GLANCE LEARNING QUESTIONS

The At a Glance table for Learning Questions provides a snapshot of what learning questions will be answered and when within the three years of implementation.

Academic Year	Fall Formative	Winter Formative	Spring Formative	Spring Annual Summative
2021-2022		Enabling Context	Enabling Context	Enabling Context
2022-2023	Measurement	Measurement	Measurement	Measurement
2023-2024	Variability in implementation	Variability in implementation	Application of Established Practice	
2024 December Final Report		Measurement		
		Variability in implementation		

AT A GLANCE OUTPUTS

The At a Glance table for Outputs provides a snapshot of what outputs will be produced and when within the three years of implementation.

Academic Year	Fall Formative	Winter Formative	Spring Formative	Spring Annual Summative
2021-2022		Enabling Context	Enabling Context	Enabling Context
2022-2023	Application of Established Practice	Measurement	Measurement	Measurement
2023-2024	Application of Established Practice	Variability in implementation		Measurement
2024 December Final Report		Variability in implementation		
		Measurement		
		Application of Established Practice		
		Enabling Context		

DETAILED TIMELINE

2021-2022 ACADEMIC YEAR

		Learning questions	Outputs
FALL FORMATIVE 2021	ENABLING CONTEXT	What conditions most distinguish districts' and schools' readiness for, approach to, and success with implementation? How have districts' priorities for implementation of new curricula or instructional systems shifted in light of COVID-19? How do these shifts influence key enabling conditions?	Definition of readiness conditions at school and district levels that support scaling effective solutions
WINTER FORMATIVE 2021	ENABLING CONTEXT	Which conditions are go/no-go factors versus important readiness conditions that a provider can help establish?	A curated list of self-assessment tools for district leaders to gauge their system's readiness and identify next steps to prepare for implementation
SPRING FORMATIVE 2022	ENABLING CONTEXT	What conditions maximize likelihood of successful implementation across contexts relevant to priority students? How does the set of key enabling conditions vary across priority contexts?	Identification of context-specific readiness factors that leaders in priority contexts should consider
SPRING ANNUAL SUMMATIVE 2022	ENABLING CONTEXT	All questions above.	Recommendations for funders, intermediaries, and districts on how to support rebuilding and/or enhancing readiness conditions that have been affected by COVID-19

2022-2023 ACADEMIC YEAR

		Learning questions	Outputs
FALL FORMATIVE 2022	Measurement	What are the best-established measures for assessing readiness to take on implementation, self-assessment of what is needed to support implementation, and factors (e.g., system conditions) that should be tracked to support implementation?	
FALL FORMATIVE 2022	APPLICATION OF ESTABLISHED PRACTICE		Description of the process district leaders use to decide whether and when to implement an evidence-based practice (e.g., based on educator workflows / jobs-to-be-done)
WINTER FORMATIVE 2022	MEASUREMENT	How do grantees use or eschew established leading and lagging measures related to both CIS and implementation to identify challenges and make improvements to systems of support? Why?	Guidance for how districts can select an implementation support provider based on core and context-specific implementation approaches
SPRING FORMATIVE 2023	MEASUREMENT	What factors not identified in established measures are critical to identifying challenges and making improvements? How do districts and school systems identify these factors?	Rating of the measurement tools according to educators' perceptions of their usability and trustworthiness
SPRING ANNUAL SUMMATIVE 2023	MEASUREMENT	All questions above.	Descriptive analysis of why district and school leaders use or eschew data in their decision-making

2023-2024 Academic Year

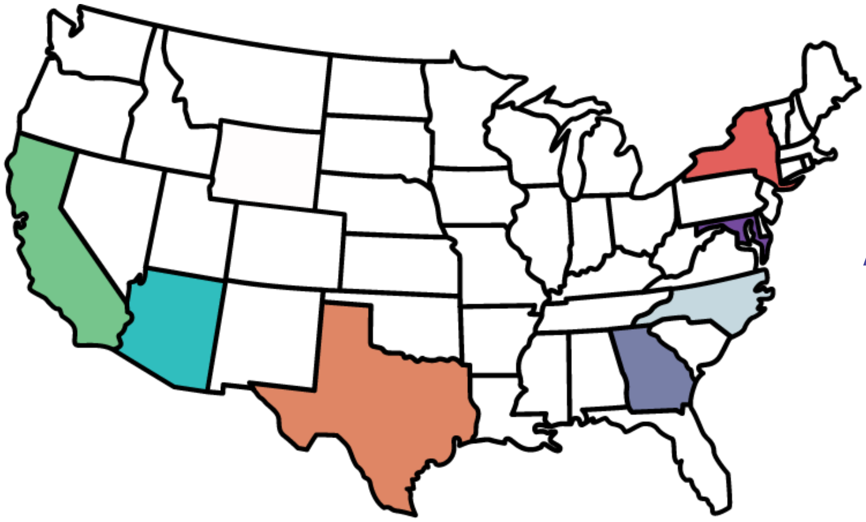
		Learning questions	Outputs
FALL FORMATIVE 2023	VARIABILITY IN IMPLEMENTATION	What is the “menu” of interventions taken by systems to support implementation? Are some seen as more promising and likely to be taken up than others? Which seem to achieve successful implementation most often? How does this vary by context?	
FALL FORMATIVE 2023	APPLICATION OF ESTABLISHED PRACTICE		An analysis of whether and how implementing a curriculum differs from implementing technical, or smaller scale solutions (e.g., a new medicine or a direct-to-student EdTech solution)
WINTER FORMATIVE 2023	VARIABILITY IN IMPLEMENTATION	What are the challenges for implementation faced across grantees? What patterns are there to those challenges, especially as related to contextual factors (e.g., student composition, characteristics of previous instructional system, teacher characteristics, broader enabling conditions)? What best practices emerge across the dimensions of implementation support that lead partners are being asked to provide, as well as any others that arise? How are those best practices related to patterns in contextual factors as described above?	Identification of core versus context-specific approaches to curriculum implementation
SPRING FORMATIVE 2024	APPLICATION OF ESTABLISHED PRACTICE	Which lessons learned from implementation research (e.g., curriculum enactment studies, education and health implementation studies) scale across this set of partnerships? Which do not? What do educators and system actors identify as the reasons why evidence-based ideas scale or do not scale?	
SPRING ANNUAL SUMMATIVE 2024	MEASUREMENT		A suite of measurement tools and new evidence to inform design and progress monitoring of solution implementation at scale

DETAILED TIMELINE

2024 December Final Report

DECEMBER 2024 FINAL REPORT	MEASUREMENT	What defines effective implementation? What are evidence-based leading indicators, lagging indicators, and data sources to assess progress and impact? How, and how much, did implementation affect student math learning and/or leading indicators of math learning? In what contexts and for whom? How do the findings from this cohort reinforce, complement, or differ from the findings about classroom-level factors that drive curricula's effectiveness in the separate enactment studies that the Foundation is funding? How does teacher perception (e.g., feasibility, acceptability, and appropriateness) of the math curriculum within CIS influence implementation?	A definition of effective implementation that districts and researchers can use to describe the strength of an implementation process
DECEMBER 2024 FINAL REPORT	VARIABILITY IN IMPLEMENTATION		A curriculum-specific model of active implementation that highlights curriculum-specific emphases such as an implementation stage based approach for curriculum

METHODS



ACTIVE PARTNERS IN:
ARIZONA
CALIFORNIA
GEORGIA
MARYLAND
NEW YORK
NORTH CAROLINA
TEXAS

SAMPLE

Participants for this investment include 11 providers and 19 Local Education Agencies (LEAs) across the United States.

Each provider has partnered with 2 to 3 LEAs for a total of 19 Provider-LEA dyads.

Criteria for inclusion in this investment included selection of a high-quality middle years mathematics curriculum, established success in providing effective professional learning services for successful implementation of the curriculum, representation of the priority student population, and ability to develop and execute data sharing agreement for each dyad.

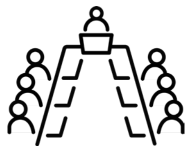
Across LEAs, six different high-quality math curricula have been identified, including: Texas Go Math (n = 3), Eureka Math (n = 1), Ready Math (n = 1), Illustrative Math (n = 8), Engage NY (n = 1), Open up Resources 6-8 (n = 1), Carnegie Learning (n = 1), Ready Classroom (n = 1) and Agile Mind (n = 2).

MEASURES

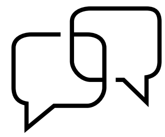
Readiness Indicators Measures

Provider-LEA pre-implementation readiness conditions will be examined across six constructs:. See **Table 1 (p.18)**. The six constructs are:

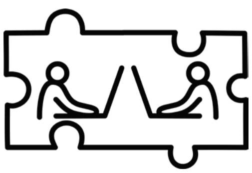
District Teaming (DT)



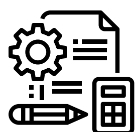
LEA Executive Sponsor Engagement (ES)



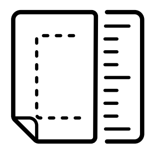
Communication (C)



Assessing Fit and Feasibility (FF)



Implementational Planning (IP)



Measurement Planning (MP)

Multiple data collection methods were used in Phase I to examine pre-implementation conditions and to inform progress monitoring. Readiness is a condition that needs to be developed and maintained throughout implementation. Given this, NIRN will continue to monitor readiness indicators and will also collect data and information from Providers and LEA partnerships in Phase II to answer the Cohort-wide learning questions.

Data will be collected through observations, electronic surveys, semi-structured interviews, listening sessions, and collection of partnership-specific learning agenda data and implementation plans.

Table 5 (p.35) depicts the data collection timeline for Phase I.

OBSERVATION MEASURES

Two observation measures will be used to collect data focused on the readiness constructs of District Teaming, LEA Executive Sponsor Engagement, Communication, Assessing Fit & Feasibility, and Implementation Planning. *Implementation Team LookFors* (Cusumano, Preston, & Ward, 2017) will be used to collect data on the quality of tools and resources developed by the Provider-LEA partnerships. A *Support Log* (NIRN, 2021) will be used to collect data on session dosage (with NIRN), topics covered, identify barriers, facilitators and successes. Data using these two observation measures will be collected at every Provider-Dyad session .

INTERVIEWS

Semi-structured interviews will be conducted with providers and participating LEA executive sponsors to collect data on the readiness constructs of District Teaming, LEA Executive Sponsor Engagement, Communication, Assessing Fit & Feasibility, and Implementation Planning. In Phase I, the semi-structured interviews were conducted by the end of May 2021.

IMPLEMENTATION TEAM (IT) SURVEY

The *IT survey* is a self-report survey which consists of 18 items. This survey will be used to collect data on the readiness constructs of District Teaming, LEA Executive Sponsor Engagement, and Communication. The survey was administered electronically through Qualtrics survey software once during Phase I to those with a team formed.

PERMANENT PRODUCTS REVIEW

A *product review rubric* was used to review the implementation plan These data will be centered on the readiness constructs of Communication, Assessing Fit & Feasibility, Implementation Planning, and Measurement Planning. Implementation specialists assigned to each Provider-LEA partnership reviewed the implementation plan at least twice in Phase 1.

TABLE 5.
Phase I Data Collection Measures and Timeline

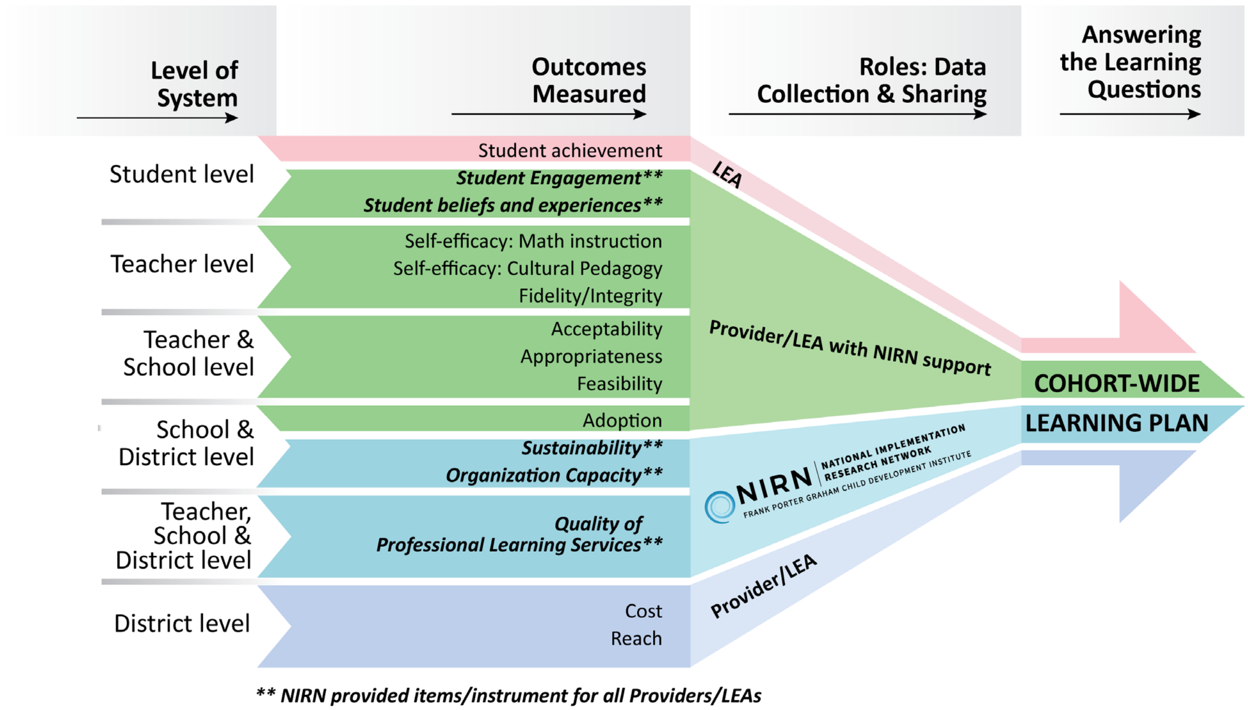
MEASURE	DATA COLLECTION METHOD	FREQUENCY & SCHEDULE	WHO WILL COLLECT?
Implementation Team Survey	Online Survey	1x (March-June)	EIC Team
Team LookFors checklist	Observation	2x (March-June)	EIC Team
Support Log	Observation	Ongoing	EIC Team
Provider Interviews	Semi-structured Interview	1x (January- March)	EIC Team
Exec. Sponsor Interviews	Semi-structured interview	1x (January- March)	EIC Team
Permanent Products Rubric	Product Review (Communication plan, implementation plan, and fit and feasibility)	1x May	EIC Team

MEASURES

Implementation, Teacher, & Student Outcome Measures

During Phase II, a number of implementation, teacher, and student outcome constructs will be examined for delivery of professional learning services and the implementation of high-quality middle years mathematics curriculum. These are organized into short-term and intermediate implementation outcomes for school and district leadership and teams, and long-term outcomes for teachers and students. The readiness indicator measures in Phase I will continue to be collected in addition to these implementation, teacher, and student outcome measures (See **Figure 3** p.36).

FIGURE 3.
EIC Measurement Plan



IMPLEMENTATION OUTCOMES (SHORT TERM & INTERMEDIATE)

Within the EIC, high-quality professional learning services and a high-quality middle years mathematics curricula are being implemented. Distinguishing implementation effectiveness from “treatment” effectiveness is necessary for translating evidence into use within classrooms. It is important to be able to determine whether lack of progress in improving outcomes for students and teachers occurred because the “treatment” was ineffective in the setting or whether the “treatment” was deployed incorrectly (i.e., implementation failure). Proctor et al. (2011) defined implementation outcomes as “the effects of deliberate and purposive actions to implement new treatments, practices, or services.” Implementation outcomes serve as indicators of implementation progress and serve as key intermediate outcomes in relation to student/teacher outcomes. Specifically, implementation outcomes are the necessary pre-conditions for obtaining desired changes in teacher and student outcomes. Conceptualizing and measuring these implementation outcomes (see Figure 3) in Phase II of the EIC will help advance understanding of implementation processes, enable comparative effectiveness of implementation strategies, and enhance efficiency in implementation research (Proctor, 2020).

To measure these implementation outcomes, a variety of methods are proposed including collecting administrative data, observations, interviews, listening sessions, and surveys.

TEACHER SURVEY

A teacher survey will be used to measure the outcomes of Acceptability, Appropriateness, and Feasibility. The survey will be administered by the Provider and/or LEA once per year in Phase II and shared annually with the Learning Partner (NIRN). The following measures have been shared with Providers and LEAs to inform their survey development.

Acceptability, or the perception of satisfaction with the professional learning services and math curriculum, will be measured using the *Acceptability of Intervention Measure* (Weiner et al, 2017). The Acceptability of Intervention Measure consists of five items on a 5-point Likert scale from "completely disagree" to "completely agree."

Appropriateness will be measured using learnings from the **BMGF's PLP investment**. Specifically, for appropriateness of professional learning services, survey items from the following scales were shared: the Professional Learning Authority Scale, Professional Learning Consistency Scale, and Professional Learning Specificity Scale. Appropriateness of the high-quality math curriculum items were shared from the Curriculum Consistency Scale and the Appropriateness of Intervention Measure (Weiner et al., 2017).

PROFESSIONAL LEARNING SPECIFICITY SCALE

measures degree of detail and clarity regarding areas such as objectives and learning goals, and connection to curriculum and state standards. It consists of six items on a 6-point Likert scale from "completely disagree" to "completely agree."

PROFESSIONAL LEARNING CONSISTENCY SCALE

measures the degree of alignment between professional learning activities, the school's mission and goals, and the district's policies. It consists of five items on a 6-point Likert scale from "completely disagree" to "completely agree."

PROFESSIONAL LEARNING AUTHORITY SCALE

measures the extent to which teachers buy into and feel supported in implementing the strategies taught in professional learning. It consists of four items on a 6-point Likert scale from "completely disagree" to "completely agree."

CURRICULUM CONSISTENCY SCALE

measures the degree of alignment between the curriculum and areas such as content standards, assessments, and school/district policies. It consists of seven items on a 6-point Likert scale from "completely disagree" to "completely agree."

APPROPRIATENESS OF INTERVENTION MEASURE

(Weiner et al., 2017) measures the perceived fit, relevance, or compatibility of the curriculum and consists of five items on a 5-point Likert scale from "completely disagree" to "completely agree."

Feasibility of the professional learning services and high-quality middle years math curriculum will be measured using the Feasibility of Intervention Measure (Weiner et al., 2017), which consists of five items on a 5-point Likert scale from "completely disagree" to "completely agree."

Quality of Professional Learning Services: Coaching will be measured through surveys Items from the Relationship with Coach scale and the Coaching Satisfaction Survey (Ihlo et al., 2017) were shared to inform survey development.

RELATIONSHIP WITH COACH SCALE

measures the degree of trust and respect established between the teacher and coach and consists of six items on a 4-point Likert scale of "not at all" to "a great extent."

COACHING SATISFACTION SURVEY

(Ihlo et al., 2017) which measures the perception of the coaching services quality and satisfaction on a 4-point Likert scale of "completely disagree" to "completely agree."

PRINCIPAL SURVEY

Principals will be surveyed by the Learning Partner (NIRN) to measure their leadership for implementation and implementation outcomes of Acceptability, Appropriateness, and Feasibility. The principal survey will be administered annually in Phase II. The survey will consist of items from the following measures: *Implementation Leadership Scale* (Aarons et al., 2014), *Acceptability of Intervention Measure*, *Feasibility of Intervention Measure*, and *Appropriateness of Intervention Measure* (Weiner et al., 2017). The survey will be administered by the Learning Partner (NIRN) once per year and will be shared with Providers and LEAs.

IMPLEMENTATION TEAM SURVEY

Members of the District Implementation Team will be surveyed by the Learner Partner (NIRN) to measure the implementation outcome of Organizational Capacity. Survey items (n = 21) will measure the functioning of the team, available leadership supports, and use of data. The survey will be administered annually by the Learning Partner (NIRN) and shared with Providers and LEAs. Note: This is the same survey that was used in Phase I but 3 additional items were included.

DISTRICT CAPACITY ASSESSMENT

The District Capacity Assessment (DCA; Ward et al., 2015) will collect data about roles, structures, functions, and systemic infrastructure to assess the district capacity to implement a high-quality middle grade math curriculum. Specifically, the DCA will collect data focused on the readiness constructs of District Teaming, Implementation Planning, and Measurement Planning.). The DCA consists of 26-items scored on a three-point scale (i.e., 2 – Fully in Place, 1 – Partially in Place, 0 – Not in Place). The assessment is completed by a trained administrator with a facilitator and team of LEA/provider participants. A total score and three subscale scores (i.e., Competency, Leadership, and Decision Support Data System) are generated. The DCA is designed to guide LEA teams to align implementation efforts and resources for the selected high-quality middle years mathematics curriculum. Psychometrically, the DCA's content validity has been established. It has an adequate internal structure (RMSEA = .071, CFI = .93, TLI = .92), internal consistency (Cronbach alphas of 0.91 for the total score and 0.79 to 0.81 for the subscale scores), and test-retest reliability (r = .98 for Leadership, .78 for Decision Support Data System and Competency Scales; Ward et al., 2020). The DCA will be administered annually by the Learning Partner (NIRN) with the district implementation team in Phase II and shared with the Provider and LEA.

INTERVIEWS AND LISTENING SESSIONS

Semi-structured interviews will be conducted by the Learning Partner (NIRN) with providers and participating LEA executive sponsors, and listening sessions will be conducted with coaches to collect data on the implementation outcomes of organizational capacity, quality of professional learning services and implementation, and sustainability. Sustainability questions will be drawn from the Program Sustainability Tool (Washington University, 2018). Interviews and listening sessions will be conducted annually in Phase II by the Learning Partner (NIRN).

ADMINISTRATIVE DATA

Administrative data will be submitted annually by the LEA to measure the outcomes of Adoption, Cost, Penetration/Reach and Sustainability. Administrative data will be shared with the Learning Partner (NIRN) annually.

Adoption and Penetration/Reach will be measured by the number of schools and teachers using High-Quality Mathematics Curriculum.

Cost will be measured by review of resources allocated on the implementation plans.

TEACHER OUTCOMES AND INSTRUCTION (INTERMEDIATE)

The intermediate teacher outcomes of self-efficacy, or degree of confidence in delivering math instruction, and using cultural pedagogy strategies, as well as fidelity/integrity of the math curriculum will be measured in Phase II (see **Figure 3** p.36) by Providers and LEAs. The data collected will then be shared with the Learner Partner (NIRN). **Table 6 (p.42)** depicts the data collection timeline for all short-term and intermediate outcomes for Phase II.

TEACHER SURVEY

A teacher survey will be used to measure the degree of confidence in delivering math instruction, meeting student needs, and using culturally pedagogy strategies. Using learnings from the **BMGF's PLP investment**, survey items will be shared with Providers-LEAs to use from Confidence in *Meeting Students Needs Scale, Culturally Responsive Teaching Self-efficacy Scale*, and *Confidence Using Ambitious Math Instruction Scale*. In addition, Providers will identify instruments to assess teacher growth in knowledge and skill in delivery of mathematics instruction aligned with the professional learning being provided. The teacher survey will be administered once a year in Phase II by the Provider or LEA and shared with the Learning Partner(NIRN) annually.

CONFIDENCE IN MEETING STUDENTS NEEDS SCALE
consists of six items on a 10-point Likert scale from 0 (not confident at all) to 10 (extremely confident).

CULTURALLY RESPONSIVE TEACHING SELF-EFFICACY SCALE
consists of eight items on a 10-point Likert scale from 0 (not confident at all) to 10 (extremely confident).

CONFIDENCE USING AMBITIOUS MATH INSTRUCTION SCALE
consists of eight items on a 10-point Likert scale from 0 (not confident at all) to 10 (extremely confident).

CLASSROOM OBSERVATIONS

Classroom observations of use of the curriculum as intended (integrity/fidelity) will be conducted by Providers and LEA staff using observation tools identified by the Provider and the LEA. Examples of observation tools proposed by Providers include those such as **MQI Observation Tool**, EQUIP tool (Reinholz and Shah, 2018), Illustrative Mathematics' IM Teams, and Classroom Observation Tool. Observation data collected by the Provider and LEA will be shared annually with the Learning Partner (NIRN).

TABLE 6.
Phase II Implementation Outcomes Data Collection, Measures and Timeline

MEASURE	DATA COLLECTION METHOD	WHO WILL COLLECT?	FREQUENCY & SCHEDULE	OUTCOME
Administrative Data Form for Reporting	Administrative Data	Provider/LEA	Annually Spring	AD, P/R, SU
District Capacity Assessment (DCA)	District Team Assessment	EIC Team	Annually Winter	OC
Implementation Team Survey	District Team Survey	Provider/LEA	Year 1: Annually Winter Year 2-3: Annually Fall	QU, OC, CO
Implementation Leadership Scale	Principal Survey	Provider/LEA	Annually - Winter	QU, OC
Measure to be selected by Providers & LEAs using surveys above (see example teacher survey)	Teacher Survey	Provider/LEA	Annually - Fall	AC, AD, AP, FE, QU, OC, TE
EIC Provider and Executive Sponsor Protocol	District team (DT) and Leadership Interviews	Learning Partner	Annually - Winter	OC, SU
EIC Coach Listening Sessions Protocol	Coach Listening Sessions	Learning Partner	Annually - Winter	QU, OC, SU
Implementation Team LookFors checklist	Observation	Learning Partner	A3 times per year (Fall, Winter, Spring)	QU, OC
Observation of Math Instruction (Classroom)	Observation	Provider/LEA	Annually - Winter	QU
Implementation Plan Review Rubric	Product Review	Learning Partner	Annually - Spring	CO

Note. AC = Acceptability; AD = Adoption; AP = Appropriateness; FE = Feasibility; QU = Quality of PL Services; OC = Organizational Capacity; CO = Cost; I/FI = Integrity/Fidelity; P/R = Penetration/Reach; SU = Sustainability

DEMOGRAPHIC DATA

District and School-level variables will be compiled to provide context for evaluation data including number of students enrolled, student-teacher ratio, percentage of students receiving free or reduced lunch, percentage of students in each racial/ethnic subgroup, and home language. (See **Appendix A** for a complete list of school, student, and teacher level demographic data). These data will be provided annually by the Provider/LEA. Socio-demographic data will be described in tables and depicted in charts.

STUDENT OUTCOMES (LONG-TERM)

Student outcome measures will be examined across the domains of student math achievement, experience in the classroom, and engagement. **Table 7 (p.44)** depicts the data collection timeline for student outcomes in Phase II.

STUDENT MATH ACHIEVEMENT MEASURES

To measure student math achievement, state and local assessments (e.g., benchmark assessments) will be used. This data will be collected by all LEAs and shared with the Learning Partner annually.

STUDENT EXPERIENCE, BELIEFS, AND ENGAGEMENT RELATED TO MATHEMATICS

To measure students' self-efficacy and confidence, value of math, enjoyment of math, and engagement with math, a student survey will be used by all LEAs. Using recommendations from **BMGF Middle Years Math Cohort**, a short student survey (n = 30 items) was developed using scales and items from the *Expectancy-Cost-Value Scale (Kosovich et al., 2014; Lauermann et al., 2017; Simpkins et al., 2006; Wang, 2012), Math and Me Survey (Adelson & McCoach, 2011), Math and Science Engagement Scales (Wang et al., 2016; Fredricks et al., 2016), National Study of Learning Mindsets (Yeager, et al., 2019)*, and *Trends in International Mathematics and Science Survey (TIMSS 2011)*. The student survey will be administered by all LEAs annually in Year 1, at Fall and Spring in Years 2-3, and shared with the Learning Partner (NIRN) annually.

TABLE 7.
Student Data Collection, Measures and Timeline

MEASURE	DATA COLLECTION METHOD?	WHO WILL COLLECT?	FREQUENCY & SCHEDULE
Adapted AMS Survey	Student Survey	Provider/LEA	Year 1: Annually-Spring Years 2-3: Fall-Spring
Student administrative records of state and local benchmark assessments	Student Administrative Records (e.g., student math achievement data,) See **Appendix A** for list of variables	Provider/LEA	Annually - Fall

MEASURES

EIC Learning Partner Effectiveness Measures

To evaluate the effectiveness of the Learning Partner's (NIRN) ability to provide high-quality services and tools, a number of formative and summative measures will be used, including: (1) professional learning evaluations to capture the quality and use of adult learning practices, knowledge gained, and follow up learning needs at every learning event; (2) consultation/coaching support effectiveness surveys and documentation logs will be monitored every three months; and (3) tracking of iterative development process from draft to finalization as compared to identified milestones and benchmarks every three months.

To evaluate the effectiveness of NIRN's ability to co-design and execute a cohort-wide learning plan inclusive of measures and baseline data, formative and summative measures will be used, including: (1) evaluation surveys to assess quality and effectiveness of facilitation and support in the co-design process every three months; (2) professional learning evaluations to capture the quality and use of adult learning practices, knowledge gained, and follow up learning needs at every learning event; and (3) formative feedback gathered throughout the process through observations.

PROCEDURES

Providers self-selected their partnerships with the LEAs to participate in this investment. LEAs are selecting the participating schools within their district during Phase I. The EIC team will hold an overview of the investment at Provider-LEA partnership sessions. The overview will include a review of the purpose, procedures, what and how data will be collected, potential risks, and their right to withdraw from the data collection activities at any time. In addition, as part of our role as a Learning Partner, cohort wide support sessions will be held quarterly to support providers and LEA teams in their use of implementation best practices and use of data for continuous improvement purposes.

This evaluation entails data collection from multiple sources, including online systems, teacher surveys, school records, and observational data collection. When possible, we will use web-based data collection to improve efficiency. **Tables 5** (p.35) and **Table 7** (p.44) above show the data collection timeline.

STUDENTS

MATH ACHIEVEMENT ASSESSMENTS

LEAs will share data from math student achievement assessments that students participate in as part of their typical educational experience. Specifically, math scale score data from state summative assessments (completed annually) and their local selected benchmark assessments (e.g., NWEA Measures of Academic Progress, i-Ready, STAR; completed three times a year) will be shared with the research team annually (Fall: October-November) for previous year assessment results.

SURVEY

LEAs/providers will share data from a web-based student survey regarding student's beliefs, engagement, and experience with math instruction. The student survey will be administered by the LEA or provider in the Spring (April-June) in Year 1, at Fall and Spring in years 2-3, and shared twice a year with the Learning Parter (NIRN).

MATH TEACHERS

SURVEY

Math teachers within participating schools will complete a web-based survey administered by their LEA or Provider partner annually (Winter: December-February in Year 1 and Fall for years 2-3) to assess their perception of the implementation of the math curriculum, professional learning services (e.g., feasibility, appropriateness, acceptability, quality), knowledge and skill in delivery of instruction, and their confidence in teaching mathematics and using culturally responsive teaching practices. The LEA or provider will share survey results annually with the Learning Partner (NIRN).

OBSERVATION

Math teachers will experience observations of their delivery of instruction by a trained observer who are either LEA (e.g., instructional coach) or Provider staff (e.g., trainer, instructional coach). Providers and LEAs will conduct observations for 30 minutes using their selected observation protocol annually in the Winter (January-February) and share the data annually with the Learning Partner (NIRN). Observers are trained by the Provider (i.e., external partner providing support to the district) and will be required to meet and maintain 80% or higher rate of agreement with the trainer in project classrooms. Reliability will be periodically tested and reported by the provider to the research team to detect observer drift and maintain consistency.

SCHOOL PRINCIPALS

SURVEY

Principals within participating schools will complete a web-based survey (*Implementation Leadership Scale*, 12 items on a 5-point Likert Scale), administered by the Learning Partner (NIRN) annually (April-June) in Year 1 and at Fall in years 2-3 to assess their leadership practices within implementation.

PROVIDER

INTERVIEWS

Semi-structured interviews will be conducted with the lead Provider for the LEA (e.g., external partner to the LEA providing professional learning supports). All interviews will be conducted annually (Winter: December-February) for 30-45 minutes by the research team. Interviews will be conducted virtually and recorded with permission from the participant for transcription and analyses.

DISTRICT STAFF

INTERVIEWS AND LISTENING SESSIONS

Semi-structured interviews will be conducted with a district executive leader (e.g., Chief Academic Officer, Assistant Superintendent). Listening sessions will also be conducted with instructional coaches providing instructional coaching to teachers, if available. All interviews will be conducted annually (Winter: December-February) for 30-45 minutes by the research team. Interviews and listening sessions will be conducted virtually and recorded with permission for transcription and analyses.

SURVEY

District staff (1) participating on the district implementation team accountable for math implementation and (2) district staff serving in the role of instructional coaches will complete a self-report web-based survey administered by the research team annually (Winter: December - February) to assess their leadership practices within implementation. The survey will be 21 questions on a 4-point Likert scale.

CAPACITY ASSESSMENT

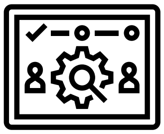
District staff participating on the district implementation team accountable for math implementation will complete a team-based district capacity assessment. The team assessment will be administered by a trained facilitator on the research team using a consensus-based scoring procedure. The assessment takes 90-120 minutes to complete. The research team will administer the team capacity assessment annually (Winter: December-February in Year 1 and September-December in Years 2-3).

OBSERVATIONS

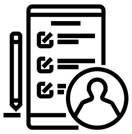
The district implementation team will be observed in their implementation team meetings by a trained observer on the research team. The trained observer from the research team will observe District team meetings annually for 30-60 minutes in length (Winter: December-February in Year 1 and at Fall, Winter and Spring in Years 2-3).

DATA ANALYSIS PLAN

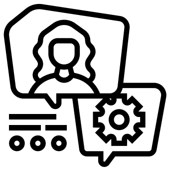
Data analyses will be conducted to answer the learning questions. We will use a mixed-method approach to address each category of questions through an analysis of qualitative and quantitative data sources. The analysis will be largely descriptive using cohort-wide trends to achieve goals including:



Creating evidenced-based models for the implementation of high-quality middle years math curricula



Adding to the knowledge base available to education professions on how to implement high-quality middle years math curricula, particularly for students of color and those experiencing poverty



Providing participating Providers and LEAs with meaningful, formative feedback to inform their practice.

The design will use the school-level as the primary unit of analysis to examine all the learning questions. Analysis of the qualitative and quantitative data will be through descriptive, visual trends, and thematic analyses. Analytically, the challenge of this investment is that there are few common measures across all schools. Within the sample, not all schools use the same high-quality math curriculum. Within a given curriculum, the sample might be smaller. Small numbers raise the risk that observed changes or trends could be due to random variation in these small groups. To the extent possible we will aggregate data across schools and LEAs. When examining math academic outcomes, we will seek to disaggregate data by race, ethnicity, home language, and income level. Finally, a control/comparison group is not being used given that the purpose of this investment is to explore implementation conditions and not establish causal relationships.

To investigate more intensively the learning question of how implementation affects student math learning, we will conduct analyses at the student- and teacher-level as the primary unit of analysis with a select number of districts. We will strive to have a subsample that is representative of key factors/variables including priority student representation and various district context factors such as locale and size.

QUALITATIVE ANALYSES

The data from all interviews will be analyzed qualitatively and summarized in stages. First, all interviews will be recorded and transcribed verbatim. After each interview, NIRN will use the implementation constructs as themes, in addition to developing themes organically as the coding occurs. We will use Thomas' (2006) general inductive approach to qualitative analysis. This approach provides researchers with a straightforward method to 1) condense raw text into a brief, summary format; 2) establish clear links between the learning questions and the summary findings derived from the raw data; and 3) develop a framework of the underlying structure of experiences or processes that emerge from the raw data. NVivo software will be used for coding and to generate themes based on the key learning questions being asked. Two members of the NIRN team will code the data to ensure inter-rater reliability. After coding separately, the coders will come together to talk about similarities and differences in codes. The themes will be synthesized into the interim and final report.

QUANTITATIVE ANALYSES

MISSING DATA

To account for missing data, we will use multiple imputations in SAS 9.4. All predictor and covariate variables will be included in the imputation procedure. Creating multiple datasets with reasonable missing data values and aggregating results from analyses using multiple datasets provide the best approximation of relationships among variables. This approach is preferred over listwise deletion or single imputation. We expect limited missing data (<1% student attrition over the year; no classroom attrition expected) but will use multiple imputation using MCMC algorithms to impute missing data (Schafer & Graham, 2002) into 40 imputation datasets. These approaches for addressing missing data are appropriate when data are missing at random (Schafer & Graham, 2002).

STATISTICAL POWER

Power analysis will be conducted to determine the minimal detectable sized effect using the dataset and ensure that the study is well powered to conduct the proposed analyses and to detect the smallest-sized effects.

PRELIMINARY ANALYSES

Preliminary factor analyses will test whether composite variables underlie our direct assessments of student's outcome measures. This will provide more precise measurement and potentially fewer tests of our constructs. For quantitative analyses using multiple measures for the same construct (e.g., teacher surveys of implementation), initial descriptives for items under each construct will be examined in order to conduct initial assessment of variability in item responses prior to psychometric modeling. For continuous items, this will include means, Pearson correlations, and standard deviations; for binary items (e.g., correct/incorrect), this will include item frequencies and tetrachoric correlations. For each construct, formalized psychometric analysis will proceed with the following steps.

First, confirmation of the factor structure (unidimensional or multidimensional, contingent on the structure of the established measure) as it has been used in past studies using a flexible form of confirmatory factor analysis (CFA) that allows for a mix of item distributions (continuous or ordered categorical).

Second, contingent on model fit, a more restrictive form of CFA model will be fit where factor loadings will be constrained to equality within each factor; this has the effect of fitting a model to test whether total scores are appropriate to use (i.e., equal weighting of items; McNeish & Wolf, 2020). The degree of misfit of this model captures the extent to which using total scores would create bias in scale score estimates and give an inaccurate picture of participant progress at the individual- and aggregate-levels.

Third, once a finalized base model is established, differential item functioning (DIF) will be tested to assess the extent to which different item parameters (i.e., item intercepts/thresholds, factor loadings) are required across different predictors of DIF such as time or demographic variables (e.g., race/ethnicity) under the moderated non-linear factor analysis (MNLFA; Bauer, 2017) framework; see also Morgan-López et al., 2020a, 2020b; Ruglass et al., 2020; Saavedra et al., 2021.

Scale scores and "personalized" standard errors of measurement that take into account a) differences in the relative weight of each item in relation to the construct and b) whether the items parameters differ across time and/or populations (i.e., free of measurement bias) would then be output from MNLFA analyses. The outputs will then be used to conduct school and student level analyses using Mplus v8.

Standardized statistical approaches will be followed to examine student math achievement data. When standardization is not feasible, harmonization will be explored as an alternative methodology. Logical harmonization involves integrating similar data elements without any data transformation. However, harmonization should involve statistical transformation, usually using Item Response Theory to bring the different data elements to the same metrics. If math assessment outcomes are statistically harmonizable, they will be integrated, and appropriate analyses will be used to create equitable scales. If instruments are not harmonizable, data-based outcomes (e.g., dichotomous variable, improvement/no improvement on study/site specific scale) will be integrated.

SCHOOL-LEVEL UNIT OF ANALYSIS

The school-level analysis will be descriptive in nature and will explore the bi-directional relationship between level of implementation support, aggregate school implementation conditions (average rates of fidelity, teacher self-efficacy, etc.), and aggregate student mathematics beliefs, experience, engagement, and achievement. While teacher and student surveys and administrative data will be collected at the individual-level, it will be aggregated to the school-level because we are not requiring a cohort-wide process of linking teacher and student data.

Through these analyses, we will be able to describe generally the extent to which provider, LEA, and school implementation conditions influence teacher self-efficacy, implementation fidelity, and aggregate student outcomes (especially for groups of priority students); and whether school-level teacher self-efficacy and implementation fidelity is related to positive student outcomes aggregated to the school-level.

STUDENT-LEVEL UNIT OF ANALYSIS

A multitude of district, school, and teacher factors influence not only how well a new curriculum is implemented, but also how students perceive and engage with mathematics and achieve on standardized state and local assessments. While more rigorous and demanding than a school-level analysis, a deeper dive with a selected group of LEAs will allow us to gain a better understanding of how implementation affects student learning, for whom, and in what contexts. In these analyses, student administrative data would be matched to teacher administrative data to control for the potential influence of teacher-level factors on student learning and achievement.

We will use propensity score analysis to ensure the group of students in the schools being included in the analysis are similar on key variables (race/ethnicity, gender, economically disadvantaged, academic achievement) as students in schools not included in the analysis. Propensity score analysis (with weights) is used to balance measured confounders or covariates that influence both the probability of selection into two or more non-experimental groups and intervention outcomes. Propensity score analysis is conducted in three steps:

1. each student's probability of receiving the treatment based on a set of covariates,
2. ensure balance between the two groups on key variables,
3. conduct multivariate analyses using the outcome(s) of interest.

We will use multilevel linear modelling to account for the nested nature of the educational data (students in a classroom within a school, within a LEA). Although there are limitations with regards to generalizability, controlling for school and teacher factors that might influence implementation fidelity and student outcomes will allow to isolate the effect of implementation on student learning, for whom, and in what contexts.

DATA MANAGEMENT

Data management is maintained in accordance with protocols and policies of the LEA and their Institutional Review Board (IRB) or Research Review Board (RRB) and other parties' IRBs including UNC-Chapel Hill's for NIRN. In addition, procedures and data sharing will be reviewed for compliance with the Family Educational Rights and Privacy Act (FERPA) and, if required, Health Insurance Portability and Accountability Act (HIPAA).

Confidentiality of data will be maintained according to the policies and procedures identified in the human subjects' protocol that will be approved by UNC-CH's IRB before beginning research activities for this project. For data use by investment staff, all electronic data will be securely stored in either the investment database or file server, maintained by UNC-Chapel Hill, and accessible only by authorized investment personnel and organization staff members responsible for managing computing equipment. Data exchanged with investment staff outside UNC-Chapel Hill will be encrypted prior to sharing.

Given the variety of data to be collected, the Provider/LEAs will submit the data via a secure file transfer protocol (SFTP) for which each Provider and LEA will be granted unique access credentials to create a separate submission portal. Data storage on the system will adhere to all laws protecting confidential information. The data team will monitor and review the data submitted to ensure high quality and completeness throughout the investment.

The first data upload will include data from the 2020-2021 school year. These data should be shared by October 30, 2021. If the data uploaded is not de-identified, the EIC team at UNC will de-identified the data and summarize before sharing. The data will be reviewed to determine if subjects could be identified deductively through small cells, and any offending demographic variables will be removed to minimize such deductive disclosure for small subsamples.
Semi-structured interview data will be recorded for accuracy and quality of data collection with the permission of participants. All data will be securely stored, and no identifiable information will be stored.

REFERENCES

Allensworth, E., Cashdollar, S., & Gwynne, J. (2021). Improvements in math instruction and student achievement through professional learning around the Common Core State Standards in Chicago. AERA Open, 7, 2332858420986872.

Althauser, K. (2015). Job-embedded professional development: Its impact on teacher self-efficacy and student performance. *Teacher Development*, *19*(2), 210-225.

Bauer D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, 22, 507–526.

Bingham, G.E., & Okagaki, L. (2012). Ethnicity and student engagement. In *Handbook of research on student engagement* (pp. 65–95). Springer.

Brock, M. E., & Carter, E. W. (2017). A meta-analysis of educator training to improve implementation of interventions for students with disabilities. *Remedial and Special Education*, 38(3), 131-144.

Bunce, A. E., Gruß, I., Davis, J. V., Cowburn, S., Cohen, D., Oakley, J., & Gold, R. (2020). Lessons learned about the effective operationalization of champions as an implementation strategy: results from a qualitative process evaluation of a pragmatic trial. *Implementation Science*, *15*(1). https://doi.org/10.1186/s13012-020-01048-1

Bryk, A. S., Gomez, L., Grunow, A., & LeMahieu, P. (2015). *Learning how to improve: How America's schools can get better at getting better*. Harvard Education Publishing.

Bui, Y. N., Fagan, Y. M. (2013). The effects of an integrated reading comprehension strategy: A culturally responsive teaching approach for fifth-grade students' reading comprehension. *Preventing School Failure: Alternative Education for Children and Youth*, 57, 59-69. doi:10.1080/1045988X.2012.664581

Cammarota, J., Romero, A. (2011). Participatory action research for high school students: Transforming policy, practice, and the personal with social justice education. *Educational Policy*, 25, 488-506. doi:10.1177/0895904810361722

Caprara, G. V., Barbaranelli, C., Steca, P., Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, 44, 473-490. doi:10.1016/j.jsp.2006.09.001

Crawford, L., Carpenter, D. M., Wilson, M. T., Schmeister, M., & McDonald, M. (2012). Testing the relation between fidelity of implementation and student outcomes in math. *Assessment for Effective Intervention*, 37(4), 224-235.

Cribbs, J., Huang, X., & Piatek-Jimenez, K. (2021). Relations of mathematics mindset, mathematics anxiety, mathematics identity, and mathematics self-efficacy to STEM career choice: A structural equation modeling approach. *School Science and Mathematics*, 121(5), 275-287.

Dee, T. S., Penner, E. K. (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal, 54*, 127-166. doi:10.3102/0002831216677002

Elaine, A. & Seamanb, J. (2017). What we teach: K-12 school district curriculum adoption process. Babson Survey Research Group. Available at https://eric.ed.gov/?id=ED582435

Evans, J. A. (2015). Gender, self-efficacy, and mathematics achievement: an analysis of fourth grade and eighth grade TIMSS data from the United States.

Finn, J. D., & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In *Handbook of research on student engagement* (pp. 97–131). Springer.

Fixsen, D., Blase, K., Horner, R., Sims, B., & Sugai, G. (2013). Scaling up brief: Readiness for change. Retrieved from: https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/SISEP-Brief3-ReadinessForChange.pdf

Flaspohler, P., Duffy, J., Wandersmann, A., Stillman, L., & Maras, M. (2008). Unpacking prevention capacity: An intersection of research to practice models and community center models. *American Journal of Community Psychology*, 41, 182-196.

Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of research on student engagement* (pp. 763–782). Springer.

Fullan, M. (2010). All systems go: The change imperative for whole system reform. Thousand Oaks, CA: Corwin.

Fullan, M. & Quinn, J. (2016). *Coherence: The right drivers in action for schools, districts, and systems.* Corwin.

George, H. P., Cox, K. E., Minch, D., & Sandomierski, T. (2018). District Practices Associated With Successful SWPBIS Implementation. *Behavioral Disorders*, 43(3), 393–406. https://doi.org/10.1177/0198742917753612

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly*, 82, 581–629. http://dx.doi.org/10.1111/j.0887-378X.2004.00325.x

Heifetz, R., & Linsky, M. (2017). *Leadership on the line, with a new preface: Staying alive through the dangers of change.* Harvard Business Press.

Higgins, M. C., Weiner, J., & Young, L. (2012). Implementation teams: A new lever for organizational change. *Journal of Organizational Behavior*, 33, 366–388.

Huang, X., Zhang, J., & Hudson, L. (2019). Impact of math self-efficacy, math anxiety, and growth mindset on math and science career interest for middle school students: the gender moderating effect. European *Journal of Psychology of Education*, 34(3), 621-640.

Hoy, A. W., Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21, 343-356. doi:10.1016/j.tate.2005.01.007

Leithwood, K., Azah, V. N. (2017). Characteristics of high-performing school districts. *Leadership and Policy in Schools*, 16, 27–53.

Levin, H. M., & Belfield, C. (2015). Guiding the development and use of cost-effectiveness analysis in education. *Journal of Research on Educational Effectiveness*, 8(3), 400-418.

Louis, K., Leithwood, K., Wahlstrom, K., & Anderson, S. (2010). Investigating the links to improved student performance. University of Minnesota. Available at https://www.wallacefoundation.org/knowledge-center/Documents/Investigating-the-Links-to-Improved-Student-Learning.pdf.

Lyon, A. R., Cook, C. R., Locke, J., Davis, C., Powell, B. J., & Waltz, T. J. (2019). Importance and feasibility of an adapted set of implementation strategies in schools. *Journal of School Psychology*, 76, 66-77.

Malen, B., Rice, J. K., Matlach, L. K., Bowsher, A., Hoyer, K. M., & Hyde, L. H. (2015). Developing organizational capacity for implementing complex education reform initiatives: Insights from a multiyear study of a teacher incentive fund program. *Educational Administration Quarterly*, 51(1), 133-176.

McIntosh, K., Mercer, S. H., Nese, R. N. T., Strickland-Cohen, M. K., Kittelman, A., Hoselton, R., Horner, R. H. (2018). Factors predicting sustained implementation of a universal behavior support framework. *Educational Researcher*, 47(5), 307–316. https://doi-org.libproxy.lib.unc.edu/10.3102/0013189X18776975

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*, 2287-2305.

Metz, A., Naoom, S. F., Halle, T., & Bartley, L. (2015). *An integrated stage-based framework for implementation of early childhood programs and systems* [OPRE Research Brief OPRE 2015- 48]. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Miech, E. J., Rattray, N. A., Flanagan, M. E., Damschroder, L., Schmid, A. A., & Damush, T. M. (2018). Inside help: An integrative review of champions in healthcare-related implementation. *SAGE Open Medicine*, *6*, 205031211877326. https://doi.org/10.1177/2050312118773261

Morgan-López, A.A., Killeen, T.K., Saavedra, L.M., Hien, D.A., Fitzpatrick, S., Ruglass, L.M., & Back, S.E. (2020) Crossover between diagnostic and empirical categorizations of full and subthreshold PTSD. *Journal of Affective Disorders*, 274, 832-840.

Morgan-López, A.A., Saavedra, L.M., Hien, D.A., Killeen, T.K., Back, S.E., Ruglass, L.M., ... & Patock-Peckham, J.A. (2020). Estimation of equable scale scores and treatment outcomes from patient-and clinician-reported PTSD measures using item response theory calibration. *Psychological Assessment, 32*, 321-335.

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., ... & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and policy in mental health and mental health services research, 38*(2), 65-76.

Redding, C. (2019). A teacher like me: A review of the effect of Student–Teacher Racial/Ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. Review of Educational Research, 89(4), 499-535. https://doi.org/10.3102/0034654319853545

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*, 700-712. doi:10.1037/a0027268

Ruglass, L.M., Morgan-López, A.A., Saavedra, L.M., Hien, D.A., Fitzpatrick, S., Killeen, T.K., Back, S.E., & López-Castro, T. (2020) Measurement non-equivalence of the Clinician-Administered PTSD Scale by race/ethnicity: Implications for quantifying PTSD severity. *Psychological Assessment*, 32, 1015-1027.

Saavedra, L.M., Morgan-López, A.A., Hien, D.A., Killeen, T.K., Back, S.E., Ruglass, L.M., Fitzpatrick, S., Lopez-Castro, T. (2021). Putting the Patient Back in Clinical Significance: Moderated Nonlinear Factor Analysis for Estimating Clinically Significant Change in Treatment for Posttraumatic Stress Disorder, *Journal of Traumatic Stress*. https://doi.org/10.1002/jts.22624

Skilling, K., Bobis, J., & Martin, A. (2015, July). The engagement of students with high and low achievement levels in mathematics. In Proceedings of the 39th Psychology of Mathematics Education conference (Vol. 4, pp. 185-192). Psychology of Mathematics Education (PME).

Sleegers, P. J., Thoonen, E. E., Oort, F. J., & Peetsma, T. T. (2014). Changing classroom practices: The role of school-wide capacity for sustainable improvement. *Journal of Educational Administration*.

Strunk KO, Marsh JA, Bush-Mecenas SC, Duque MR. The Best Laid Plans: An Examination of School Plan Quality and Implementation in a School Improvement Initiative. Educational Administration Quarterly. 2016;52(2):259-309. doi:10.1177/0013161X15616864

Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation, 27*(2), 237-246. Doi: 10.1177/1098214005283748

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., ... & Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. Nature, 573(7774), 364-369.

Wandersman, A., & Scaccia, J. (2018). Readiness briefing paper. Retrieved from: https://www.wandersmancenter.org/uploads/1/2/8/5/128593635/531c7f_a2db100ea8d04e8b8dceea2c696bc855.pdf

Ward, C., St. Martin, K., Horner, R., Duda, M., Ingram-West, K., Tedesco, M., Putnam, D., Buenrostro, M., & Chaparro, E. (2015). District Capacity Assessment. University of North Carolina at Chapel Hill.

Warwick, J. (2008). Mathematical self-efficacy and student engagement in the mathematics classroom. *MSOR Connections*, 8(3), 31-37.

APPENDIX A

The data being requested will include records for all middle school students enrolled at any time in the given year requested and attending schools participating in the EIC project. While the EIC is focused on students who are experiencing poverty, Black, Latino/a, and/or English Learner (EL)-Designated ("priority students"), it is critical to have data on all students from all participating schools to address the learning questions of the investment. **All measures should be reported by student once a year.** A basic description of the variables to be requested appears in the table below.

All student data should be shared for the following academic years: 2021-2022, 2022-2023, 2023-2024.

Teacher data should be shared for the following academic years: 2021-2022, 2022-2023, 2023-2024.

ADMINISTRATIVE DATA

CATEGORY	VARIABLE	DETAILS
Local Identifiers	Unique Student ID number	A stable, unique identifier for each student that allows data for an individual student to be linked across all files and submissions *See the section below regarding the Study ID.
Enrollment	School Name	
	Grade Level	
Demographics	Gender	For systems that do not have mid-year grade-level changes, once per year is acceptable
	Race	All possible categories, codes, and information about how classifications are applied are needed (i.e., how handle more than one race code, race/ethnicity combined, etc.)
	Ethnicity	Hispanic or Latino Not Hispanic or Latino
	HomeLang	As reported by families. ***If coded differently please provide codebook**
	State defined economically disadvantaged status	
	Designated as limited English proficient (LEP)	
	Special education status	

ADMINISTRATIVE DATA

CATEGORY	VARIABLE	DETAILS
Math Course Information	Teacher	A variable with the teacher's name or identifier for each teacher. (optional)
	Math course name	
	Math Course attendance	
	Math course final grade	
Attendance* (if available)	Number of Absences	
	Number of Days Enrolled	
Math Assessment Information	State Assessment Scale Score – Mathematics	
	State Assessment Proficiency Level Score – Mathematics (if available)	
	Local Benchmark – Mathematics Assessment Name	Indication of which assessment the student took (e.g. grade level, name)
	Local Benchmark Mathematics Results	For the assessment taken, indication of achievement, reported as a level of proficiency, scaled or raw score (in order of preference).

ADDITIONAL DATA FOR STUDENTS BEING COLLECTED FOR LEARNING AGENDA TO BE SHARED

CATEGORY	VARIABLE	DETAILS
EIC Student Survey	EIC Student Survey Data of Beliefs, Engagement, and Experience in Mathematics	All raw data

TEACHER ADMINISTRATIVE DATA

CATEGORY	VARIABLE*	DETAILS
Teaching Site	School Name	
Local Identifier	Unique Teacher Identifier	A stable, unique identifier for each teacher that allows data for an individual teacher to be linked across all files and submissions with student data
Demographics	T-Reside	How long has the teacher resided in the United States.
	Ethnicity	Hispanic or Latino Not Hispanic or Latino
	Race	NIRN will need all possible categories, codes, and information about how classifications are applied (how they handle more than one race code, are race and ethnicity combined, etc.)
	Education	
	National Board Certified	
	Years as a teacher	

ADDITIONAL DATA FOR TEACHERS BEING COLLECTED FOR LEARNING AGENDA TO BE SHARED

CATEGORY	VARIABLE	DETAILS
EIC Teacher Survey(s)	Teacher survey(s) of self-efficacy in mathematic and culturally responsive teaching, as well as implementation constructs (e.g., acceptability, feasibility)	All raw data
Classroom Observation	Teacher observations of mathematics instruction using identified observation tool	Raw data for fidelity/integrity observations as well as the criteria to determine if fidelity/integrity was met Raw data for fidelity/integrity observations as well as the criteria to determine if fidelity/integrity was met

APPENDIX B

The Effective Implementation Cohort (EIC) Readiness Constructs and indicators are implementation best-practices that are shown to facilitate the uptake of evidence-based practices. This document defines the following constructs in terms of observable and measurable indicators: District Teaming, LEA Executive Sponsors, Communication, Fit and Feasibility Assessment, Implementation Planning, and Measurement Planning. Feedback from participating EIC cohort members (providers and district staff) and the Bill & Melinda Gates Foundation's National Math Advisory Council have been incorporated into the readiness constructs' definitions and indicators.

EIC READINESS CONSTRUCTS AND DEFINITIONS

READINESS CONSTRUCT	CONSTRUCT DEFINITION
District Teaming	A representative team has been formed at the district level and is working to lead implementation and create the enabling conditions in selected schools so that educators can make full use of high-quality middle-years math curriculum.
LEA Executive Sponsor Engagement	The Executive Sponsor(s) champions and supports district and school staff as they engage in implementation of high-quality middle years math curriculum. An executive sponsor is an individual with the ability to influence others, authority to make decisions regarding resource allocation, institutional knowledge, time, and positive relationships with staff and critical perspectives.
Communication	Frequent and accurate information regarding implementation planning and progress is exchanged between critical perspectives and acted upon by the identified persons. Some goals of communication include sharing information, gathering feedback and input, clarifying expectations, and celebrating successes.
Assessing Fit and Feasibility	An assessment conducted by districts to better understand how a new or existing high-quality middle years mathematics curriculum works within their existing context to support implementation planning and use.
Implementation Planning	Comprehensive implementation strategies are specified within a plan to ensure capacity is developed to support successful use and sustainability of a high-quality middle years mathematics curriculum.
Measurement Planning	Utilize multiple methods to collect and review data to: (1) inform decision-making for continuous improvement, (2) examine effectiveness, and (3) communicate with critical perspectives.

DISTRICT TEAMING

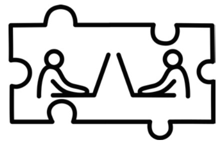
Construct Definition	A representative team has been formed at the district level and is working to lead implementation and create the enabling conditions in selected schools so that educators can make full use of high-quality middle-years math curriculum.
Indicators	District Team is representative of the diversity of the district, schools, and priority students being served and the necessary math content knowledge, leadership/decision-making authority, cross-functional perspectives from district and schools, and provider support
	District Team utilizes effective team meeting and communication processes and procedures
	Attendance at meetings is high (greater than 80% of team members) and consistent
	District Team members have sufficient time dedicated to planning for and leading implementation functions (in and outside of meetings) or fulfilling their role and responsibilities
	Positive working relationships with building/campus leadership, school/campus teams, and staff
	District Team members follow through and accomplish identified actions within needed timelines
	District Team has clear vision, mission, and shares accountability for the work
Optional Resources	**District Implementation Team Handout** **Guiding Questions for Team Development** **Active Implementation Hub Implementation Teams Learning Module**

LEA EXECUTIVE SPONSOR ENGAGEMENT

Construct Definition	The Executive Sponsor(s) champions and supports district and school staff as they engage in implementation of high-quality middle years math curriculum. An executive sponsor is an individual with the ability to influence others, authority to make decisions regarding resource allocation, institutional knowledge, time, and positive relationships with staff and critical perspectives.
Indicators	Lead Partner and district team cultivate the role of LEA executive sponsor(s) and school leaders as champions
	Executive sponsor(s) and Lead Partner collaborate to support an effective partnership
	Executive sponsor(s) use data and information for planning and/or problem solving
	Executive sponsor(s) use effective bi-directional communication with district leadership, team, staff, and stakeholders (e.g., board, community)
	Executive sponsor(s) ensure that priority students are centered when making decisions for prioritizing the work, equitably allocating resources, and addressing implementation barriers
	Executive sponsor(s) can speak to and answer questions about implementation of the high-quality middle years math curriculum and motivate staff for implementation (visible promotion)
	Executive sponsor(s) is an individual with decision-making authority
Optional Resources	**Executive Sponsor Handout** **Active Implementation Hub Leadership Learning Module**

COMMUNICATION

Construct Definition	Frequent and accurate information regarding implementation planning and progress is exchanged between critical perspectives and acted upon by the identified persons. Some goals of communication include sharing information, gathering feedback and input, clarifying expectations, and celebrating successes.
Indicators	District Team identifies internal and external stakeholders for bi-directional communication
	District Team plans for communication messages, methods, and frequency for different stakeholders including school leadership, teams, and staff to celebrate successes and be responsive to identified needs in planning and implementation
	District Team uses coherent communication messages and methods/tools as outlined in their plan
	Effective and efficient communication processes are used between Provider, District Team, and schools
	District Team assesses effectiveness of communication using feedback from stakeholders
Optional Resources	**Active Implementation Hub Linking Communications Protocol Lesson** **Linking Communication Protocol Handout**



ASSESSING FIT AND FEASIBILITY

Construct Definition	An assessment conducted by districts to better understand how a new or existing high-quality middle years mathematics curriculum works within their existing context to support implementation planning and use.
Indicators	Key stakeholders (e.g., school staff, students, families) are involved in identification of the needed changes related to mathematics for the priority students
	Assessment considers evidence, usability, available supports for the mathematics curriculum, how well it addresses the district's needs related to mathematics for priority students, the fit with priorities and values of the district, schools, families, and community, the district's previous experience with implementing new practices or changes, and the capacity of the district to sustain and scale implementation
	Assessment identifies initiatives with potential competing resources, capacity, and identifies mitigating factors
	Indication of assessment of policies and procedures or impact analysis related to math (e.g., access to rigorous courses, etc.) for priority students
Optional Resources	**Hexagon Tool: Fit & Feasibility Assessment** **Active Implementation Hub Hexagon Tool Process Learning Module** **Initiative Inventory** **Active Implementation Hub Initiative Inventory Learning Module**

IMPLEMENTATION PLANNING

Construct Definition	Comprehensive implementation strategies are specified within a plan to ensure capacity is developed to support successful use and sustainability of a high-quality middle years mathematics curriculum.
Indicators	Provider/LEA define shared Theory of Action to improve outcomes for priority students
	Activities are specified/identified to develop staff competency (e.g., professional learning, coaching) around the use of the chosen mathematics curriculum
	Provider and LEA demonstrate shared accountability for co-creation and use of implementation strategies
	Priority students are kept at the center of implementation planning and activities (e.g., asking who benefits from decisions and who may be further marginalized and what will this team do to mitigate this gap; how will priority students be engaged)
	Data are used to identify needs, set goals and implementation benchmarks, and monitor progress
	Provider and LEA incorporate strategies that ensure equitable resource allocation
	Planning activities progress amid competing priorities
Optional Resources	**Implementation Stages Planning Tool** **Active Implementation Hub Implementation Stages Planning Tool Learning Lesson** **Active Implementation Hub Implementation Stages Learning Module** **Active Implementation Hub Implementation Drivers Learning Module**

MEASUREMENT

Construct Definition	Utilize multiple methods to collect and review data to: (1) inform decision-making for continuous improvement, (2) examine effectiveness, and (3) communicate with critical perspectives.
Indicators	Measurement plan includes process, programmatic, and impact measures for identified implementation goals and benchmarks
	Multiple methods and sources are used to collect data (e.g., observation, surveys, interviews/focus groups, assessments, administrative data) to support having appropriate and relevant evidence for decision making
	Indicators are well defined: leading/lagging, how data will be collected, who will collect/compile necessary data, and how frequently it will be collected
	Process exists for team to review, problem-solve and develop an action plan using disaggregated data as it relates to priority students and share results with stakeholders
Optional Resources	**Active Implementation Hub Decision Support Data System Learning Lesson** **Active Implementation Hub Fidelity Learning Module**