Implementation Development Map: Literature Review for Formative Child Assessment

October 2021

Megan Hague Angus, Rebecca Kleinman, Jessica F. Harding

Submitted to:

Bill & Melinda Gates Foundation
P.O. Box 23350
Seattle, WA 98102
Phone: (206) 709-3100
Program Officer: Sarah Weber

Submitted by:

Mathematica
P.O. Box 2393
Princeton, NJ 08543-2393
Phone: (609) 799-3535
Project Director: Jessica F. Harding

This page has been left blank for double-sided copying.

Literature Review for Formative Child Assessment (FCA)

In this document, we summarize our literature review on the Formative Child Assessment (FCA) element of the Implementation Development Map (IDM). We start with an overview, then provide a bulleted list that summarizes the strength of support from professional/expert recommendations or the research literature, and we discuss whether the research speaks to equity. Following the detailed notes are two graphics that summarize, for each IDM indicator, the strength of (1) the research evidence and (2) the support from expert recommendations and professional best practices. The appendix describes our literature search and review process.

A. Overview

Formative Child Assessment (FCA) is one of seven elements in the IDM. FCA focuses on establishing policies, resources, and infrastructure to support the collection of child assessment data at the state and local levels. FCA examines the types of child assessment data collected and the extent to which states and programs use those data to set goals and establish action plans. As with other elements of the IDM, the FCA element has two types of indicators: those about infrastructure at the state level (five indicators) and those about implementation at the local level (seven indicators). Infrastructure indicators cover statewide formative child assessment activities, such as creating policies and guidelines for conducting formative child assessments, distributing resources, offering professional development, and engaging families in assessments. Implementation indicators look at programs' use of formative child assessments, to improve classroom quality and inform professional development and supports for teachers who implement them.

Definitions

Research strength is based on the number of *high quality studies* with favorable effects on child or teacher outcomes.

- High quality studies are those in which the design is strong enough to suggest that outcomes can be attributed to the intervention, practice, or policy that is being studied.

Practice strength is based on whether the indicator is supported by professional best *practices or expert* recommendations.

- Professional best practice standards include the Head Start Performance Program Standards (HSPPS) and the standards set forth by the National Association for the Education of Young Children (NAEYC).
- Expert recommendations are from the National Academy of Sciences, Engineering, and Mathematics (NASEM).

At the request of the Bill & Melinda Gates Foundation, Mathematica conducted a systematic literature review focused on FCA. (The full methodology appears in the appendix.) For the FCA element, after screening the studies collected for the literature review, we identified and reviewed three studies published since 2001 to assess their quality and key findings (see References). For this element, one study supported at least one IDM indicator. Despite the limited availability of high quality research, we caution readers against drawing conclusions about the inherent value of an IDM indicator. Readers should not conclude that a lack of high quality studies means that the indicator does not have valuable, nuanced information to offer about how to strengthen state systems.

Because the IDM is a tool designed to improve state systems, we also determined which elements and indicators were supported by professional best practice standards and expert recommendations. (The box on the first page defines high quality, best practice standards, and expert recommendations; see the

appendix for full definitions and a description of how we rated these dimensions to determine the overall research strength and practice strength of each IDM indicator.)

The IDM tool explicitly embeds equity into the indicators to ensure state leaders continue to value diverse groups of learners and teachers and provides high quality learning opportunities for all children. In our literature review, we examined equity by describing and placing value on studies that include students and teachers with diverse characteristics. We have captured whether the samples in high quality studies with favorable effects include dual language learners (DLLs), children whose families have low incomes, and children and teachers of racially and ethnically diverse backgrounds. Research that explicitly addresses questions of equity is limited, however, despite its importance for state systems that serve children from disadvantaged backgrounds.

B. Details of support for indicators

In this section, we describe the strength of support the indicators have from the research literature or the recommendations of professionals and experts. We detail the high quality studies with favorable effects, the parts of the indicator supported by the study, and any themes in the results that concern outcomes of children or teachers. We report whether any studies are particularly relevant to a specific IDM indicator and whether the research addresses equity, particularly whether studies were based on diverse samples or showed effects for certain groups of children or teachers. If there are no studies related to an indicator (Figures 2 and 4), we do not discuss it.

IDM FCA 1: State requires programs to use formative child assessments (FCAs) that include all three of the following characteristics:

- **Valid and reliable**
- **Comprehensive across all domains and development**
- **Aligned to state learning and development standards**

Practice strength

- Both the professional and expert recommendations support this indicator. Both sets supported the idea of using valid, reliable child assessment measures to support early learning, suggesting that teachers and other staff should use the information collected from valid and reliable assessments to evaluate a child's development and progress.

IDM FCA 2: State requires programs to have a process in place so that FCA data on children who are Dual Language Learners DLLs and children with developmental delays and disabilities is valid and not misinterpreted due to language or cultural barriers. The state provides clear guidelines on how to do this (e.g., using interpreters or linguistically appropriate assessments) and verifies directly or indirectly that programs comply.

Practice strength

- Both the professional and expert recommendations support this indicator. Both sets note that formative child assessments should be appropriate for DLLs and children with special needs. Both sets of recommendations note that the assessments should be suitable, valid, and reliable for the

population and purpose for which they are to be used. Neither speak specifically about the role the state should play in providing clear guidelines on verifying programs are in compliance.

IDM FCA 3. State provides resources (e.g. funding, guidance, etc.) to support teachers in the implementation of formative child assessments.

Practice strength

- Both the professional and expert recommendations support this indicator. The recommendations state that programs should provide training and guidance to teachers in how to assess children and then use the information collected by formative child assessments. NASEM specifically notes that policymakers, program directors, assessment administrators, and practitioners should have access to ongoing opportunities to understand assessments and use the collected data for the proper purposes.

IDM FCA 4: State verifies that programs collect FCA data and programs use it to:

- **Inform improvement plans**
- **Track progress**
- **Communicate data to families.**

In addition, state uses the data to guide decisions for technical assistance and resource allocation to programs.

Practice strength

- Both the professional and expert recommendations support this indicator. Both suggest that formative child assessments should be used to support individualized learning, improve teaching practices, and enhance overall knowledge of a child's development. Both sets of recommendations also suggest that parents and families are and should be a key audience for assessment information.

IDM FCA 5: The state's efforts to understand and address inequity include ongoing data collection, disaggregation of data, active discussions, data-driven decision-making, action planning, implementing, assessing implementation, and refining as needed. The state specifically collects data to understand and address the following four components:

- **FCAs are conducted in reliable and valid methods for all child populations**
- **Data are used to inform instruction, especially instruction that supports the learning and development of children from dual language and special needs backgrounds**
- **All teachers have access to trainings on fidelity of implementation of assessment instruments, as well as trainings on bias when collecting and interpreting data**
- **Teachers involve parents in child's education by sharing formative assessment data and seeking guidance from parents when creating individualized instruction plans**

Practice strength

- Neither the professional nor expert recommendations support this indicator. They do not specifically address how states or programs should understand and address inequity through the disaggregation of data.

IDM FCA 6. Programs use valid and reliable formative child assessment measures that are comprehensive across all domains and aligned to state learning and development standards.

Practice strength

- The professional recommendations partially support this indicator. The HSPPS and NAEYC both assert the value of a programs' use of valid and reliable formative child assessment measures. They do not suggest that the assessments should be aligned with state standards. The expert recommendations support this indicator; they suggest that the assessments should be aligned with standards.

IDM FCA 7: Programs have a process in place so that formative child assessment data of children from special populations are valid and accurate.

Practice strength

- Both sets of professional and expert recommendations support this indicator. Both sets acknowledge that children from special populations (DLLs and children with special needs) have the right to be assessed. The recommendations underscore the importance of appropriately and accurately assessing children who are DLLs and children with special needs.

IDM FCA 8. Programs use formative child assessment data for classroom quality improvement.

Research strength

- One study examined the effects of administering formative child assessments to preschool children and the relationship between progress monitoring and student performance. This randomized trial found that regularly measuring and graphing children's literacy skill development resulted in positive effects on student achievement compared to children in the control group.
- Progress monitoring was effective for racially/ethnically diverse children. Over 50 percent of the children in the study were identified as non-White.

Practice strength

- Both sets of professional and expert recommendations support the use of formative child assessment data to improve teaching practices. Data can be used for planning activities, for tracking learning individually to assess a child's strengths or needs, and for informing classroom instruction.

IDM FCA 9. Programs use formative child assessment data to plan professional development and CQI work.

Practice strength

- Both sets of professional and expert recommendations support this indicator. They note that the use of formative child assessment data for professional development purposes and for continuously improving programming. NASEM notes that those working in early childhood classrooms and programs should be purposeful in their educational planning and should use assessments for planning and monitoring what children are learning.

IDM FCA 10. Teachers are trained and, when appropriate, certified in the FCAs they implement.

Research strength

- One study examined the effects of administering formative child assessments on preschool children and the relationship between progress monitoring and student performance. This randomized trial trained teachers to measure children's literacy skill development and to use a program to graph children's results. A randomly assigned subset of teachers also received additional training and consultation. Teachers who were trained to use formative child assessments had students with higher achievement compared to a control group of teachers who did not administer or receive training in formative child assessments. But student achievement did not differ for the teachers who received any training compared to the teachers who received additional consultation.
- Training teachers in progress monitoring was found to be effective for racially/ethnically diverse children. Over 50 percent of the children in the study were identified as non-White.

Practice strength

- Both sets of professional and expert recommendations support this indicator. The professional and expert recommendations underscore the importance of ensuring that qualified staff members assess children. NAEYC recommends that staff should be familiar with and knowledgeable about assessment and that programs should ensure that in-service training builds teachers' and administrators' literacy around assessments. NASEM also states that teachers and other staff must receive training in and follow-up on the use of assessment tools.

IDM FCA 11. Programs have supports (e.g., written materials, in-person or online training, teacher collaboration groups) for FCA implementation available to teachers.

Research strength

- One study examined the effects of administering formative child assessments on preschool children and the relationship between progress monitoring and student performance. This randomized trial trained teachers to measure children's literacy skill development and to use a program to graph children's results. A randomly assigned subset of teachers also received additional training and consultation. Teachers who were trained to use formative child assessments had students with higher achievement compared to a control group of teachers who did not administer or receive training in formative child assessments. But student achievement did not differ for the teachers who received any training compared to the teachers who received additional consultation.
- Progress monitoring was effective for racially/ethnically diverse children. Over 50 percent of the children in the study were identified as non-White.

Practice strength

- Both sets of professional and expert recommendations support this indicator. The professional recommendations suggest that programs should make additional resources available to teachers to support formative child assessments. NAEYC notes that staff should be given resources to support their knowledge and skills about early childhood assessment.

IDM FCA 12. Teachers regularly share children's progress with families in culturally and linguistically sensitive ways and in their preferred language.

Practice strength

- Professional recommendations partially support this indicator. They suggest that teachers should engage and communicate with parents, but the recommendations do not explicitly suggest that teachers should share information in the children's and families' preferred language. Expert recommendations support this indicator; they suggest that teachers should share information with and involve parents in culturally and linguistically inclusive ways. NASEM notes that families with DLLs should play critical roles in the assessment process. It argues that the educational benefit of an assessment is optimized when DLL parents' voices are included and considered throughout the process.

C. Overall rating of research and practice support for indicators

Figures 1 and 2 summarize the overall strength of the research and practice support for each FCA indicator.

Figure 1. Indicator key for overall ratings of research and practice strength

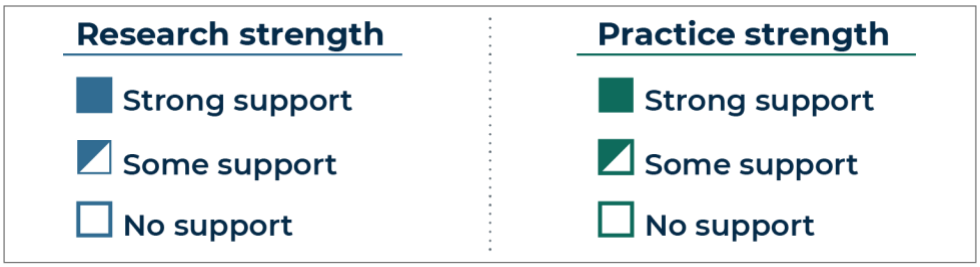
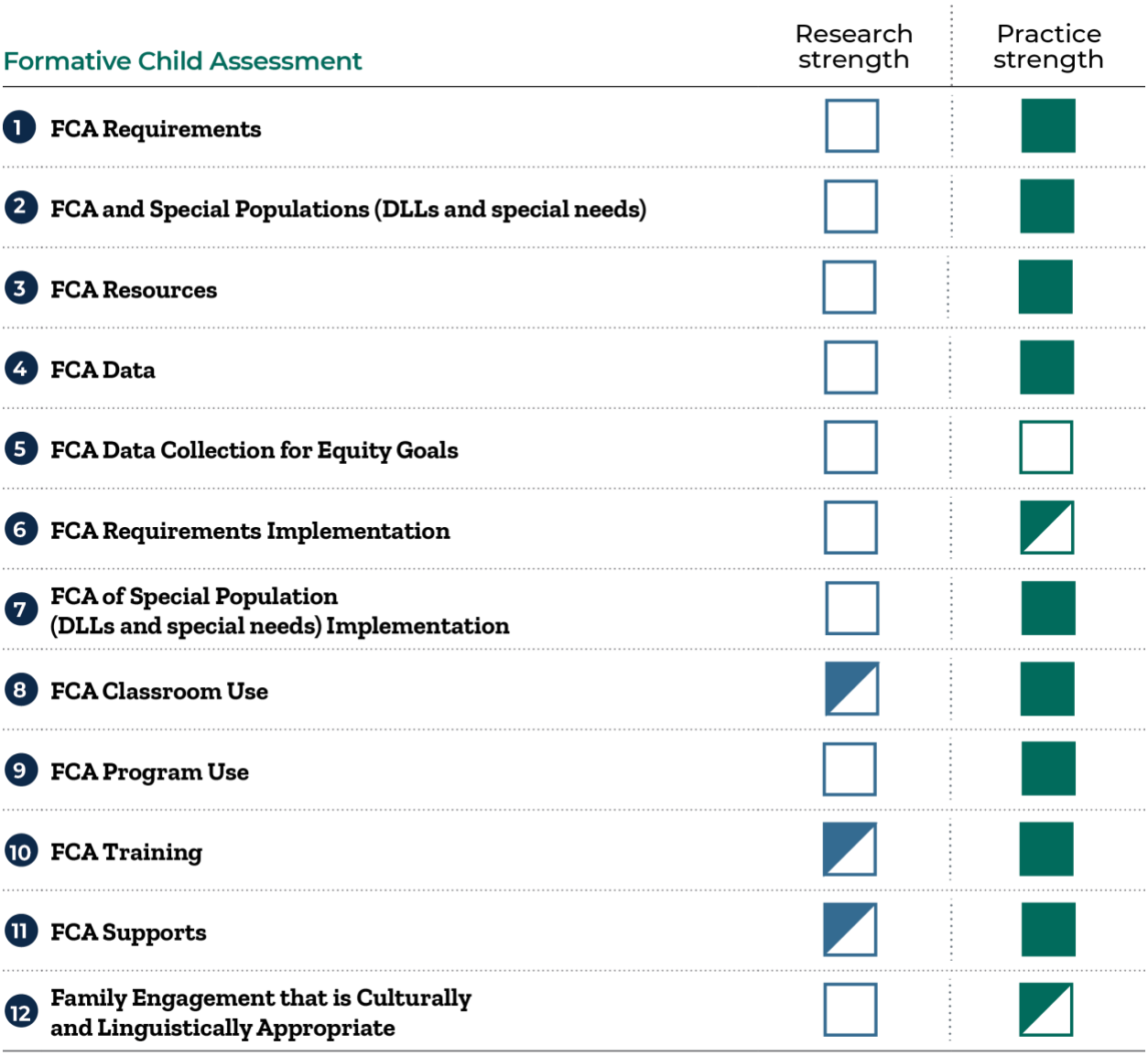


Figure 2. Overall ratings of research and practice strength



D. Detailed ratings of research and practice support for indicators

Figures 3 and 4 give additional detail on the research and practice support for each IDM indicator.

Figure 3. Indicator key for detailed ratings of research and practice strength

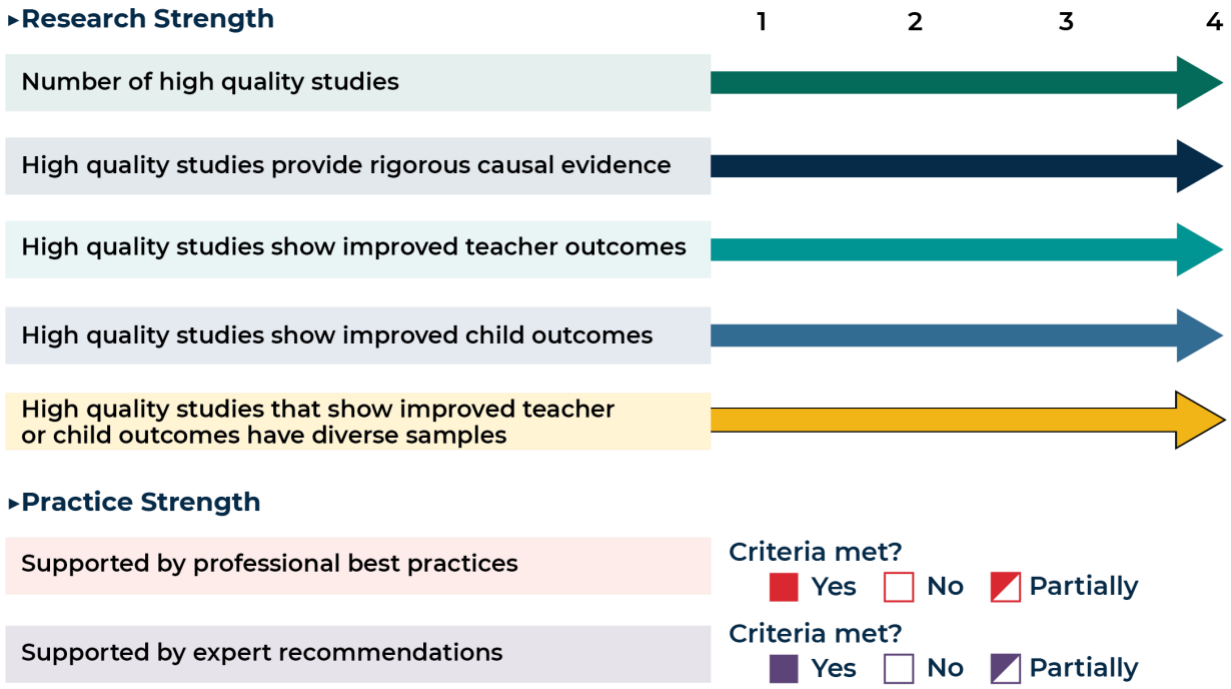
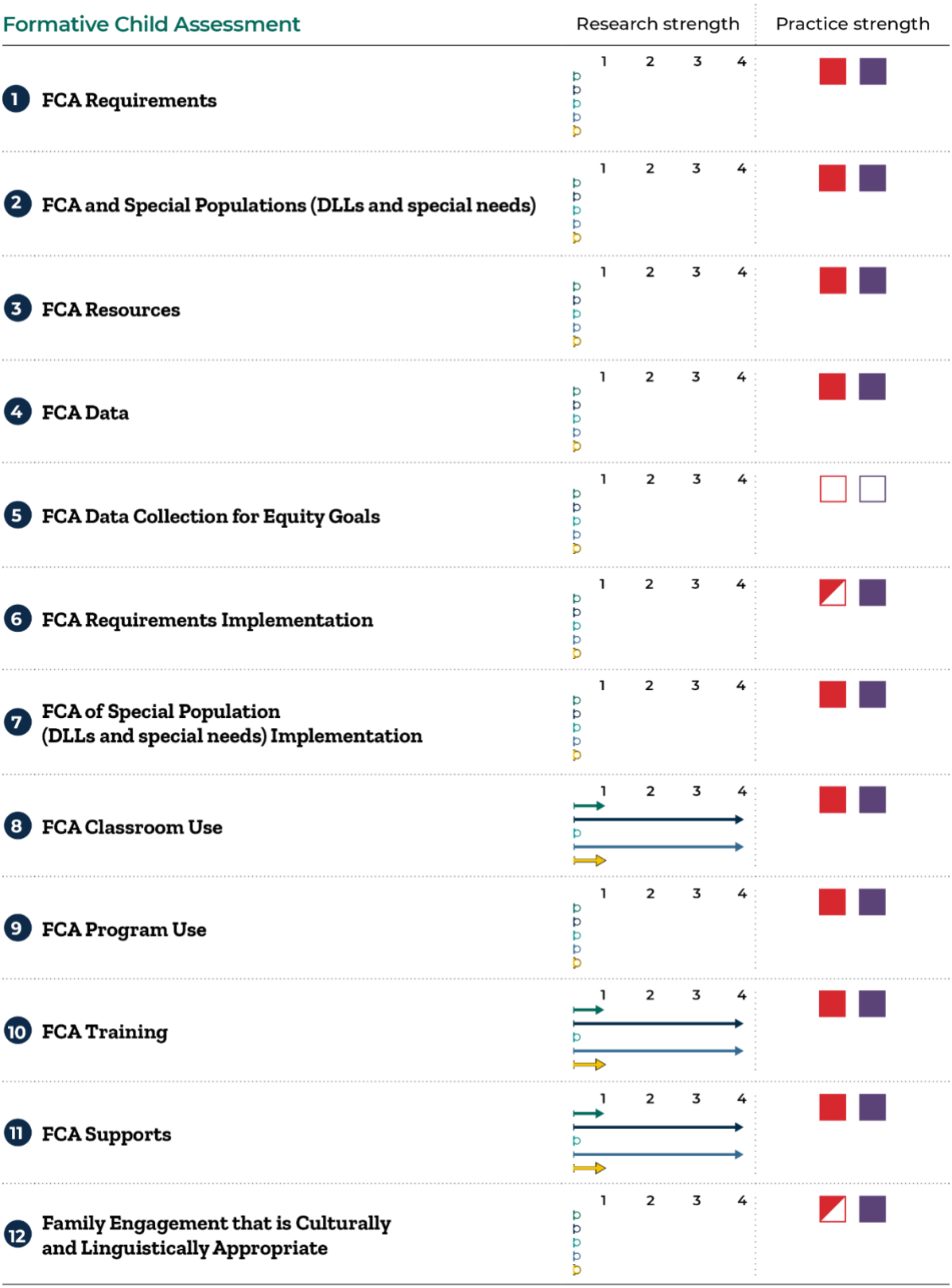


Figure 4. Detailed ratings of research and practice strength



Appendix

A. Identifying literature

Mathematica staff reviewed the literature on state-level FCA policies and guidelines and the use of formative child assessments to improve instruction in preschool classrooms. We worked with our professional librarians to develop targeted search terms. We then searched eight databases for published articles.[1] Using the information in the abstracts, we screened out studies that did not meet our inclusion criteria. All eligible studies had to meet the following criteria:

- Based in the United States
- Focused on children ages 3 to 5
- Implemented in a prekindergarten setting (Head Start, child care center, or state prekindergarten program)
- Evaluated child or teacher/classroom outcomes using a randomized controlled trial, quasi-experimental, or correlational design
- Published in 2001 or later

We procured the full text of the eligible studies. Next, we screened the studies again to identify whether the studies mapped to any of the Implementation Development Map (IDM) indicators and to confirm that the studies met our inclusion criteria. We screened out any studies that did not focus on an IDM indicator (Table A.1). For the FCA element, after examining the full text of the 13 studies initially identified, 3 met the inclusion criteria and were rated high quality, and 1 was rated high quality and had at least one favorable outcome (see the reference list for the high quality studies).

Table A.1. Number of studies identified, reviewed, and found to support the FCA element

IDM element	Studies identified	Studies fully reviewed	High quality studies	High quality studies with favorable outcomes
Formative child assessment	13	3	3	1

B. Assessing support for IDM indicators

We assessed each indicator on seven dimensions (Tables A.4 and A.5) to summarize the support for the indicator in the research and professional/expert recommendations.

To identify high quality studies, reviewers rated the rigor of the study design (Dimensions 1 and 2). To identify whether the studies show an improvement in outcomes, reviewers summarized the study impacts on children or teachers (Dimensions 3 and 4). To identify the extent to which high quality studies provided evidence of improvements with diverse groups of children and teachers, reviewers examined the groups of children and teachers included in the studies (Dimension 5). To determine the extent to which professional best practices and expert recommendations supported the indicators, we reviewed key practice documents (Dimensions 6 and 7). Below, we describe each step.

[1] The eight databases are Academic Search Premier, APA PsycInfo, Cochrane Database of Systematic Reviews, Education Research Complete, ERIC, ProQuest Dissertations, SAGE Journals, and Scopus.

1. Rating study quality

We wanted to identify studies with results we could be confident were valid. We categorized studies as those that provide rigorous causal evidence, strong evidence, or low quality evidence (Table A.2).

Table A.2. Study quality ratings

Study rating	Description
Provides rigorous causal evidence[a]	Well-conducted randomized controlled trials with limited attrition (< 20 percent) and no other design concerns provide the strongest evidence because outcomes can be attributed to the intervention, practice, or policy rather than to existing differences between groups.
Provides strong evidence[a]	Studies that show that their comparison groups are similar or include relevant control variables suggest that outcomes can be attributed to the intervention, practice, or policy but that unmeasured differences might exist between groups. These studies could include randomized controlled trials with high attrition or quasi-experimental designs that (a) show that the comparison groups used in analysis were equivalent on demographics and a baseline measure of the outcome (or another outcome in the same domain) or (b) controls for demographics and baseline measures. These studies could also include correlational designs and ones that have a comparison group but no baseline measures, provided they use a strong set of relevant controls (including demographics and other characteristics that could influence the outcome).
Provides low quality evidence	These are studies with unconvincing results. These studies could include randomized controlled trials with high attrition, quasi-experimental designs, or correlational studies that do not use adequate control variables or that have a confound such as using different data collection methods in the treatment and comparison groups.

[a] Both of these ratings were considered to provide high quality evidence.

We then summarized the number of high quality studies—studies that provide rigorous causal evidence and strong evidence—and the percentage of high quality studies that provide rigorous causal evidence for each indicator. Studies can support several indicators.

2. Rating study findings

We categorized whether the high quality studies had statistically significant effects on any child or teacher/classroom outcomes included in the studies (Table A.3).

Table A.3. Definitions of study impacts

Study impacts	Definition
Favorable	Significant effects on at least one outcome that benefits children or teachers/classrooms; for example, improving classroom quality
Unfavorable	Significant negative effects on at least one outcome for children or teachers/classrooms and no favorable effects on any outcomes; for example, children's receptive vocabulary scores decrease
No effect	No significant effects on any child or teacher/classroom outcomes
Mixed	At least one favorable and unfavorable effect

We next summarized for each indicator the percentage of high quality studies with favorable effects on children, teachers/classrooms, or both.

3. Rating whether studies include diverse samples

For high quality studies with favorable effects on children and teachers/classrooms, we examined whether the studies included different population groups. We assessed whether studies reported that they included the following:

- Racially/ethnically diverse children (at least 25 percent of children are Hispanic, African American, or American Indian/Alaska Native)
- Racially/ethnically diverse teachers (at least 25 percent of teachers are Hispanic, African American, or American Indian/Alaska Native)
- Children who are dual language learners (DLLs) (at least 25 percent of children are DLLs)
- Children from low-income households (at least 75 percent of children are in low-income households or the educational setting is low income)

We then looked at whether each indicator has high quality studies with favorable effects with racially/ethnically diverse children, racially/ethnically diverse teachers, DLLs, and children from low-income households.

4. Assessing professional best practices and expert recommendations

Because the IDM is a tool designed to improve state systems, we determined which elements and indicators were supported by professional best practice standards, including the Head Start Performance Program Standards, the standards set by the National Association for the Education of Young Children, and expert recommendations from the National Academy of Sciences, Engineering, and Mathematics. The latter organization analyzes available evidence to advance the learning and development of children, youth, and families and presents consensus recommendations that undergo peer review before publication.[2]

A team of researchers reviewed IDM indicators to determine how well they aligned or agreed with these professional standards. We assessed whether each indicator was supported by professional recommendations and expert recommendations by using a three-part scale that included "met," "partially met," or "not met." We used "partially met" when aspects of the indicator were supported, but not necessarily when the full indicator was met, because each indicator often covers several ideas.

5. Assigning overall ratings on dimensions

Based on the rating of study quality, study findings, the diversity of samples, and professional and expert recommendations, we rated each indicator on seven dimensions (Table A.4 and Table A.5). Ratings for

[2] These documents included (1) HSPPS as found online: https://eclkc.ohs.acf.hhs.gov/policy/45-cfr-chap-xiii/1302-33-child-screenings-assessments; (2) National Association for the Education of Young Children, "Screening and Assessment of Young English-Language Learners," Supplement to the NAEYC and NAECS/SDE Joint Position Statement on Early Childhood Curriculum, Assessment, and Program Evaluation (Washington, DC: NAEYC, November 2005); (3) Early Childhood Curriculum, Assessment, and Program Evaluation, "Building an Effective, Accountable System in Programs for Children Birth through Age 8," Joint Position Statement of the National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education (NAECS/SDE) (Washington, DC: NAYEC, November 2003); (4) National Research Council, "Early Childhood Assessment: Why, What, and How" (Washington, DC: National Academies Press, 2008).

the research support dimensions ranged from 1 to 4; ratings for the recommendation support dimensions included met, partially met, and not met.

Table A.4. Definitions of dimension ratings for research support

Research support dimension	1	2	3	4
Number of high quality studies	1 to 3 high quality studies	4 to 6 high quality studies	7 to 9 high quality studies	10 or more high quality studies
High quality studies that provide rigorous causal evidence	1–25% of high quality studies provide causal evidence	26–50% of high quality studies provide causal evidence	51–75% of high quality studies provide causal evidence	76–100% of high quality studies provide causal evidence
High quality studies that show improved teacher/classroom outcomes (show at least one favorable effect on a teacher outcome and no unfavorable effects)	1–25% of high quality studies show improved teacher/classroom outcomes	26–50% of high quality studies show improved teacher/classroom outcomes	51–75% of high quality studies show improved teacher/classroom outcomes	76–100% of high quality studies show improved teacher/classroom outcomes
High quality studies that show improved child outcomes (show at least one favorable effect on a child outcome and no unfavorable effects)	1–25% of high quality studies show improved child outcomes	26–50% of high quality studies show improved child outcomes	51–75% of high quality studies show improved child outcomes	76–100% of high quality studies show improved child outcomes
High quality studies that show improved teacher or child outcomes with diverse samples	Studies include one of the following groups: racially/ethnically diverse children, racially/ethnically diverse teachers, DLLs, children from low-income households	Studies include two of the following groups: racially/ethnically diverse children, racially/ethnically diverse teachers, DLLs, children from low-income households	Studies include three of the following groups: racially/ethnically diverse children, racially/ethnically diverse teachers, DLLs, children from low-income households	Studies include four of the following groups: racially/ethnically diverse children, racially/ethnically diverse teachers, DLLs, children from low-income households

DLLs = dual language learners.

Table A.5. Definitions of dimension ratings for practice support

Practice support dimension	Not met	Partially met	Met
Supported by professional best practices	The indicator was not supported by the HSPPS or NAEYC	Part of the indicator was supported by the HSPPS or NAEYC	The full indicator was supported by the HSPPS or NAEYC
Supported by expert recommendations	The indicator was not supported by NASEM	Part of the indicator was supported by NASEM	The full indicator was supported by NASEM

NAEYC = National Association for the Education of Young Children; NASEM = National Academies of Sciences, Engineering, and Medicine; HSPPS = Head Start Program Performance Standards.

6. Assigning overall ratings on research and practice strength

To make the recommendation support rating even more accessible, we summarized two dimensions of support: research strength and practice strength (Table A.6).

Table A.6. Definitions of research and practice strength ratings

Recommendation support dimension	No support	Some support	Full support
Research strength (number of high quality studies with favorable effects on child or teacher/classroom outcomes)	No high quality studies show improved child or teacher/classroom outcomes	One or two high quality studies show improved child or teacher/classroom outcomes	Three or more high quality studies show improved child or teacher/classroom outcomes
Practice strength (whether supported by professional best practices or expert recommendations)	Neither professional best practices nor expert recommendations support the indicator	At least one set of professional best practices or expert recommendations partially supports the indicator, or only one (and not both) set fully supports the indicator	Both professional best practices AND expert recommendations support the indicator

References

Buysse, V., Peisner-Feinberg, E., Soukakou, E., Fettig, A., Schaaf, J., & Burchinal, M. Using Recognition & Response (R&R) to improve children's language and literacy skills: Findings from two studies. *Early Childhood Research Quarterly, 36*(3), 11–20.

DeLoach, D. (2012). Effects of a prekindergarten mathematics intervention on mathematical abilities of preschoolers with low socioeconomic status [Doctoral dissertation., Walden University]. ProQuest LLC. https://eric.ed.gov/?id=ED551913.

Hollman, A. K. (2009). The effects of progress monitoring and consultation on emergent literacy performance as measured by the individual growth and development indicators [Doctoral dissertation, University of Minnesota]. University of Minnesota Digital Conservancy. https://conservancy.umn.edu/handle/11299/54208.