# Developing a Tool to Examine Teachers' Use of Ongoing Child Assessment to Individualize Instruction

OPRE Report 2015-87
September 2015

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

# Developing a Tool to Examine Teachers' Use of Ongoing Child Assessment to Individualize Instruction

OPRE Report 2015-87

September 2015

Shannon Monahan, Sally Atkins-Burnett, Barbara A. Wasik, Lauren Akers, Felicia Hurwitz, and Judith Carta

**Contract Number:** HHSP23320095642WC/HHSP23337041T

**Mathematica Reference Number:** 40158

**Submitted to:**
Laura Hoard and Nina Philipsen Hetzner, Project Officers
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services
370 L'Enfant Promenade, S.W.
Washington, DC 20447

**Submitted by:**
Shannon Monahan, Project Director
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
Telephone: (609) 799-3535
Facsimile: (609) 799-0005

This report is in the public domain. Permission to reproduce is not necessary.

**Suggested citation:**

Monahan, S., S. Atkins-Burnett, B. A. Wasik, L. Akers, F. Hurwitz, and J. Carta. "Developing a Tool to Examine Teachers' Use of Ongoing Child Assessment to Individualize Instruction." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation, 2015.

**Disclaimer**

The views expressed in this publication do not necessarily reflect the views or policies of the Office of Planning, Research and Evaluation, the Administration for Children and Families, or the U.S. Department of Health and Human Services.

MATHEMATICA Policy Research

TEMPLE UNIVERSITY College of Education

KU

OPRE

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

## ACKNOWLEDGMENTS

The authors wish to thank Laura Hoard and Jennifer Brooks, our federal project officers at the Office of Planning, Research, and Evaluation (OPRE), as well as OPRE Society for Research in Child Development Fellows Nina Philipsen Hetzner and Kelly Fisher, for their oversight, guidance, and vision. We are also grateful for the ongoing input of Amanda Bryans from the Office of Head Start (OHS). Other staff members from OHS and OPRE provided us with valuable feedback about the needs of the national Head Start program and the research community. In addition, we were fortunate to have the expertise of our expert panel members, whose guidance on the topic of ongoing assessment and individualization and insights on the measure were vital.

We are very grateful to the providers who allowed us into their settings and met with us as we developed and refined the measure. We learned a great deal from these experiences.

Finally, we gratefully acknowledge our colleagues at Mathematica Policy Research who contributed their time, effort, and expertise to this project. These staff include Kimberly Boller, Susan Sprachman, Emily Moiduddin, and Nikki Aikens, who provided thoughtful quality assurance; and Patricia Ciaccio who provided editorial support. Laura Sarnoski designed the figures in the report. Karen Groesbeck formatted this report.

While acknowledging the contributions of these individuals, the authors take responsibility for any errors or omissions that remain. The views expressed in this report do not necessarily reflect those of the Administration for Children and Families or the U.S. Department of Health and Human Services.

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

## EXPERT PANEL MEMBERS

We would like to thank the following members of our expert panel for their assistance with this project. The views expressed in this publication do not necessarily reflect the views of these members.

**Stephen Bagnato**
University of Pittsburgh

**Linda Broyles**
Southeast Kansas Community Action Program

**Virginia Buysse**
FPG Child Development Institute

**Lynn Fuchs**
Vanderbilt University

**Leslie Nabors Oláh**
Educational Testing Service

**Sheila Smith**
Columbia University

**Patricia Snyder**
University of Florida

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

## CONTENTS

ACKNOWLEDGMENTS ..... III

EXPERT PANEL MEMBERS ..... V

OVERVIEW ..... XI

EXECUTIVE SUMMARY ..... XIII

I. INTRODUCTION ..... 1

A. Justification for the measure ..... 1

B. Literature review ..... 2

C. Conceptual model that informs the EDIT ..... 5

1. Stage 1: Selecting the target and assessment method ..... 6
2. Stage 2: Implementing ongoing assessment ..... 6
3. Stage 3: Interpreting data and formulating instructional decisions ..... 7
4. Stage 4: Applying instructional decisions and individualizing ..... 7

D. Contextual, teacher, and family factors affecting most stages ..... 7

E. Measurement model for the EDIT ..... 8

F. Road map for the report ..... 9

II. THE EXAMINING DATA INFORMING TEACHING (EDIT) MEASURE: INSTRUMENT AND ADMINISTRATION ..... 11

A. Rationale for a multi-method measure focused on two domains ..... 11

1. The need for a multi-method approach ..... 11
2. Measurement issues during the development phase ..... 13
3. Initially targeting two domains ..... 14

B. Overview of the three methods ..... 15

1. Document review ..... 16
2. Video-based classroom observations ..... 16
3. Teacher interview ..... 17

C. Summary of the measure ..... 18

1. Number and types of items ..... 18

D. Administration procedures ..... 20

1. Selecting focal children ..... 20
2. Teacher activities before the raters' visit ..... 20
3. EDIT rater(s)' visit ..... 21

III. THE MEASUREMENT DEVELOPMENT PROCESS ..... 25
A. Recruitment ..... 25
B. Pretest timeline and data collection procedures ..... 26
1. Recruitment and consent ..... 26
2. Selecting focal children ..... 27
3. Teacher activities before the raters' visit ..... 28
4. Raters' visit ..... 30
5. Self-administered questionnaire ..... 30
6. Teacher debrief ..... 30
C. Refining the instrument and its procedures ..... 30
IV. PRETEST ..... 31
A. Sample: Who participated in the pretest? ..... 31
1. Teachers ..... 31
2. Communication with families ..... 32
3. Classroom assessment systems ..... 32
B. How well does the EDIT capture a range of quality in practices in pretest classrooms implementing ongoing assessment and individualizing? ..... 33
1. Items with high mean scores ..... 34
2. Evidence of a need for additions or changes to the measure ..... 35
V. LESSONS LEARNED FROM THE PRETEST ..... 39
A. Teacher debrief information: How did teachers feel about participation? ..... 39
B. What were challenges and changes to EDIT procedures? ..... 40
1. Selection of focal children ..... 40
2. Describing and accessing documentation ..... 41
3. Time period for documentation ..... 42
4. Focus of video recording ..... 43
C. What are the implications of the pretest for the feasibility of the measure if brought to scale and recommendations for future procedures? ..... 43
D. What were the contributions of different data sources? ..... 44
E. How did the EDIT items change during the pretest? ..... 45
1. Items with limited variance ..... 45
2. Identified need to increase the EDIT's measurement of teachers' planning and evaluation of progress ..... 47
3. Use of "don't know" ratings ..... 50
4. Summary ..... 51

VI. FUTURE OF THE EXAMINING DATA INFORMING TEACHING MEASURE ..... 53

A. Proposed plan for further pretesting ..... 53

1. Plans for continued pretesting ..... 54

2. Benefits of continued pretesting ..... 54

B. Potential future uses ..... 55

1. Research ..... 55

2. Professional development ..... 55

3. Evaluating and informing how programs support teachers in conducting and using assessment data ..... 56

REFERENCES ..... 57

## APPENDICES

APPENDIX A: TEACHER SELF-ADMINISTERED QUESTIONNAIRE (SAQ)

APPENDIX B: LETTER TO TEACHERS

APPENDIX C: TRAINING PLAN

APPENDIX D: ADDITIONAL METHODS CONSIDERED

## TABLES

I.1 Designs of the studies identified by the literature review ..... 3

II.1 Number and types of items, by conceptual model stage ..... 19

II.2 Proposed schedule for site visit for one part-day classroom ..... 21

III.1 Rounds of data collection for the pretest ..... 25

III.2 Depiction of differences in pretest data collection, by round ..... 26

III.3 Consent collection during pretest ..... 27

IV.1 Two ongoing assessment systems ..... 32

V.1 Estimated burden on teachers ..... 39

V.2 Examples of changes made to procedures ..... 40

V.3 Changes for items with limited or no variance ..... 47

V.4 Revisions made to the EDIT to increase measurement of teacher's intentionality ..... 48

V.5 Additional EDIT items ..... 50

## EXHIBITS

II.1 Examples of documentation ..... 16

II.2 Definitions of EDIT item types ..... 18

IV.1 EDIT team reflections on nine teachers' use of assessment ..... 37

## FIGURES

I.1 Conceptual model for curriculum-embedded approaches ..... 6

II.1 Multi-method measure model ..... 15

II.2 Method for the focus and frequency of video recordings ..... 21

II.3 Use of methods to assign scores ..... 22

III.1 Alternative methods for the focus and frequency of video recordings ..... 29

VI.1 Continued EDIT pretest activities ..... 54

## OVERVIEW

In 2012, the Office of Planning, Research, and Evaluation at the Administration for Children and Families (ACF) engaged Mathematica Policy Research and its partners to conduct a project titled "Assessing Early Childhood Teachers' Use of Child Progress Monitoring to Individualize Teaching Practices." This report describes the iterative development of the Examining Data Informing Teaching (EDIT) measure. The EDIT is specifically designed to help researchers understand how teachers conduct ongoing assessments for individualization and use those assessments to guide instruction. The measure focuses on the processes the teacher uses for (1) planning what information to collect and how to do so, (2) collecting valid data, (3) organizing and interpreting the data, and (4) using the data collected to inform both overall and individualized instruction. The EDIT uses a multi-method approach in gathering evidence with checklists, ratings, and rubrics that describe how the teacher collects and uses assessment. EDIT raters review assessment and instructional planning documents, as well as video recordings of assessments and instruction. Raters also conduct a one-hour individual teacher interview.

Teachers in nine classrooms were selected to participate in the pretest. Supervisors or research partners nominated teachers who collected and used ongoing assessment information regularly. Overall, there was limited variability in teachers and classrooms receiving high ratings on many of the initial EDIT items. We found evidence of high scores on some general assessment practices. For example, teachers regularly documented information about children objectively, and they collected the documentation during meaningful[1] and authentic classroom activities. Because only nine teachers were in the pretest, it is not clear how prevalent these behaviors might be among other early childhood teachers. It does suggest that these will be among the easiest items on the EDIT.

We also identified some items that were more difficult and allowed us to examine the lower and middle ends of the rubrics, but did not allow us to examine the high end of the rubrics. In addition, we made some observations that suggested we were not measuring some constructs well on the early versions of the EDIT. This was particularly the case for constructs related to how intentionally teachers gathered and used assessment for monitoring progress and selecting instructional strategies.

Throughout the pretest, we revised our data collection and administration procedures to ensure that we were able to capture the variation in teacher practice and validly represent how teachers use assessment. We changed the order of the questions and added prompts in the teacher interview to better elicit evidence of teachers' planning and evaluation of progress. We refined the wording on the rubrics to clarify and specify concepts, and added indicators to better capture measurement of teachers' intentional planning and evaluation of progress. To further refine the EDIT and test the psychometric properties of the revised version, we recommend visiting additional classrooms with a wider range of assessment practices, including greater diversity in the level of teacher skill and in the type of assessment system used. Eventually, researchers and professional developers could use the EDIT to help design programs to support ongoing assessment practices in classrooms that could ultimately improve children's school readiness.

[1] Meaningful activities support learning in the goals and objectives being assessed.

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

## EXECUTIVE SUMMARY

In 2012, the Office of Planning, Research, and Evaluation at the Administration for Children and Families (ACF) engaged Mathematica Policy Research and its partners to conduct a project titled "Assessing Early Childhood Teachers' Use of Child Progress Monitoring to Individualize Teaching Practices." The purpose of the project was twofold: (1) to develop a research-informed conceptual model for early childhood teachers' use of ongoing assessment to individualize instruction, and (2) to create a measure to examine this process. Prior reports describe in detail the results of a literature review, conceptual framework, and measurement plan (Akers et al. 2014; Atkins-Burnett et al. 2014). This report describes the iterative development of the Examining Data Informing Teaching (EDIT) measure.[2] This report includes the results of a pilot study in nine classrooms and a proposal for next steps for the EDIT.

The EDIT is designed to examine how a teacher conducts ongoing assessments for individualization and uses those assessments to guide instruction. The focus is on the processes the teacher uses for (1) planning what information to collect and how to do so, (2) collecting valid data, (3) organizing and interpreting the data, and (4) using the data collected to inform both overall and individualized instruction. The EDIT uses a multi-method approach in gathering evidence with checklists, ratings, and rubrics that describe how the teacher collects and uses assessment. EDIT raters review assessment and instructional planning documents, as well as video recordings of assessments and instruction. Raters also conduct a one-hour individual teacher interview to probe for additional explanations about the documents and observations, as well as to obtain information on the teacher's planning and implementation of instructional adaptations, modifications, and individualized teaching strategies.

The study involved three rounds of iterative testing for the EDIT measure. After each round, in consultation with ACF, the EDIT development team met and discussed challenges that arose and proposed changes to improve the procedures and EDIT items to obtain stronger measurement. We then incorporated recommended changes, and repeated the data collection and review process in the next set of classrooms as we iteratively refined the EDIT. Between the second and third round of data collection, we shared the EDIT with the project's expert panel. The final round of data collection, in fall 2014, incorporated feedback from the expert panel.

Teachers in nine classrooms[3] were purposively selected to participate in the pretest. Supervisors or research partners nominated teachers who collected and used ongoing assessment information regularly. Seven of the nine teachers had at least a bachelor's degree, and two of those teachers also had master's degrees. The teachers' experience in their early childhood classrooms ranged from 10 months to 16 years. Six classrooms used Teaching Strategies GOLD (TS GOLD; Teaching Strategies, Inc. 2011), and the remaining three classrooms used the Work Sampling System (WSS; Meisels et al. 2001). The teachers had at least some training, as well as coaches or mentors to support the use of these assessment systems.

Teachers in the pretest sample consistency received high scores on several EDIT items. We found evidence of some high quality general assessment practices that suggests it is relatively

[2] The EDIT was formerly called the Tool for Tailored Teaching (T3).

[3] Budget (and OMB policy) restrictions limited the sample size to nine classrooms.

easy for teachers to score highly with certain types of items. For example, teachers regularly documented information about children objectively, and they collected the documentation during meaningful activities. The electronic ongoing assessment systems (e.g., TS GOLD) provided structured output that organized the data to facilitate interpretation. Most teachers planned individualized lessons, using curricular materials aligned with the objectives that they assessed. These types of practices were indicative of the behavioral descriptions anchoring the high end of items and did not offer the opportunity to evaluate how well the rubrics worked for the middle of the scale to its lower end.

We also identified some items that were more difficult and allowed us to examine the lower end and the middle of the rubrics, but did not allow us to examine the high end (for example, planning for assessments, collecting more than one piece of evidence for ratings). In addition, we made some observations that suggested we were not measuring some constructs well on the early versions of the EDIT. This was particularly the case for constructs related to how intentionally teachers gathered and used assessment for monitoring progress and selecting instructional strategies. The following observations suggest that we need to examine how we measure assessment practices related to intentionality of assessment and focus on progress over time in addition to performance as a snapshot in time. Reflection on our observations in relation to the scores on the early version of the EDIT suggest that we were not adequately capturing some important constructs:

- **Intentionality.** Evidence of a child's skill, knowledge, or behavior was limited in most areas to one or two observational records, and the method for collecting these records often was inefficient. The early versions of the EDIT did not capture the intentionality in data collection.
- **Instructional strategies.** Evidence of teacher awareness of how to identify evidence-based or professionally recommended instructional strategies was limited to the use of curricular materials. The strength of the link between child performance and the curricular activity was not reflected in the early versions of the EDIT.
- **Performance and progress.** When asked about using data for instructional decisions, teachers talked about performance rather than progress. They discussed where the children should be by the end of the year, but none of the teachers could tell us how much progress children should be making in a given month or reporting period toward that end-of-year goal.
- **Planning and intentionality.** Evidence of planning and intentionality in collecting and using data to inform instruction and individualization was not adequately captured in the initial set of items.
- **Continued monitoring.** The initial set of items did not adequately capture continued monitoring of individualized goals or the review of the effectiveness of any instructional strategy or intervention. Teachers did not describe using assessment data to determine whether an instructional approach was helping or not.

Throughout the pretest, we revised our data collection and administration procedures to ensure that we were able to capture the variation in teacher practice and validly represent how teachers use assessment. We changed the order of the questions and added prompts in the teacher interview to better obtain evidence of teachers' planning and evaluation of progress. We refined

the wording on the rubrics to clarify and specify concepts, and added indicators to better capture measurement of teachers' intentional planning and evaluation of progress. We retained all three data sources (document review, video recorded observations, and teacher interview). Each contributed some independent and some overlapping evidence about how teachers implement the process of using assessment to inform instruction and individualization. To further refine the EDIT, we recommend visiting additional classrooms with a wider range of assessment practices, including classrooms that are implementing a standard tool (e.g., General Outcome Measures) in addition to curriculum-embedded assessments. In these classrooms, we would seek to refine our items related to teacher intentionality in planning and using assessment, interpreting child progress, and use of information to inform next steps in both instruction and assessment. In addition, we could gather samples of assessment data to use in creating training materials for the EDIT.

The EDIT is specifically designed to help researchers understand how teachers use ongoing assessment to inform instruction. After additional development and further research, the goal is for the measure to be used to develop professional development materials that will improve the quality of ongoing assessment practices in classrooms and ultimately improve children's school readiness.

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

# I. INTRODUCTION

In 2012, the Office of Planning, Research, and Evaluation at the Administration for Children and Families (ACF) engaged Mathematica Policy Research and its partners to conduct a project titled "Assessing Early Childhood Teachers' Use of Child Progress Monitoring to Individualize Teaching Practices." The purpose of the project was twofold: (1) to develop a research-informed conceptual model for early childhood teachers' use of ongoing assessment to individualize instruction, and (2) to create a measure to examine this process. The result of this endeavor is the development of the Examining Data Informing Teaching (EDIT) measure. The EDIT is a measure designed to examine how a teacher conducts ongoing assessments for individualization and uses those assessments to guide instruction, regardless of the ongoing assessment tool used. In other words, the EDIT will rate the quality of a teacher's use of the ongoing assessment system, rather than the quality of the system itself. The EDIT is specifically designed to help researchers understand how teachers use ongoing assessment to inform instruction. Eventually, researchers and technical assistance providers might use the EDIT to help support the use of ongoing assessment practices in classrooms.

This chapter of the report provides (1) an overview of the justification for the measure, (2) a brief description of a literature review (including key findings), (3) a summary of the conceptual model that guided the development of the EDIT, and (4) a measurement model for the EDIT. The chapter ends with a road map for the rest of the report.

## A. Justification for the measure

Assessment has long played a critical role in helping researchers, education professionals, and policymakers examine whether early education promotes children's readiness for school. For many years, child assessment was most often used for summative purposes: to provide information on children's developmental status at different times and to show how they performed relative to peers or to specified criteria. Recent policies, however, have brought a rising interest in how teachers use ongoing assessment to adjust their teaching to best meet each child's needs. In fact, individualized teaching is a requirement in the Head Start Performance Standards (*Federal Register* 2011).

One common form of ongoing assessment is "progress monitoring." This is a scientifically based practice that assesses children's performance in a variety of domains and uses child data to inform, measure, and modify instructional practices (National Center on Student Progress Monitoring 2012). A recent wide-scale application of progress monitoring is to use response to intervention (RTI)—an approach to early intervention, sometimes called "recognition and response," that includes the regular screening of all children throughout the year (Buysse and Peisner-Feinberg 2013; Hamilton et al. 2009; National Association for the Education of Young Children et al. 2012). In RTI systems, children identified as achieving lower than average progress receive more intensive instruction, as well as frequent monitoring to gauge the success of the instructional interventions.

Several experimental studies have shown progress monitoring to be an important part of successful teacher support or professional development programs (Buzhardt et al. 2011, 2010; Landry et al. 2011). Research demonstrates that teachers who are supported in using ongoing

assessment to individualize their instruction design stronger, more effective instructional programs, and have students who achieve better outcomes, than teachers who do not assess progress (Connor et al. 2009; Fuchs et al. 1984). The use of ongoing assessment data—often merged with other professional development supports, such as mentoring—is also linked to growth in literacy outcomes in preschool through first grade (Ball and Gettinger 2009; Landry et al. 2009; Wasik et al. 2009).

Head Start recognizes the importance of using ongoing assessment to individualize instruction for each child. Over the past five years, the Office of Head Start (OHS) has elaborated on its vision for preschool child and family outcomes, strengthened its focus on monitoring program and classroom quality, and developed tools to support ongoing assessment in daily practice (U.S. Department of Health and Human Services 2010; Atkins-Burnett et al. 2009). Currently, all Head Start Centers are required to implement some form of assessment to monitor children's progress and to collect individual child information three times a year (Administration for Children and Families 2011).[4]

Despite the importance of using ongoing assessment data to guide instruction, and the Head Start program requirements to do so, information is sparse on how early education teachers actually collect and use these data to tailor their instruction. Policymakers, practitioners, and researchers continue to see an urgent need for research in this area in the quest for better educational outcomes (Bambrick-Santoyo 2010; Buysse and Peisner-Feinberg 2013; Fuchs and Fuchs 2006; Hamilton et al. 2009; Marsh et al. 2006). To determine whether teachers are implementing ongoing assessments as intended and using the data from the assessments to inform instruction tailored to children's individual needs and skills, a measure is needed to assess teacher implementation and use of ongoing assessment. To develop a measure, it is important to review the literature to understand how early childhood teachers use ongoing assessment to individualize instruction.

## B. Literature review

In 2012, the study team conducted a structured literature review to inform the development of a conceptual model and measurement plan to assess whether and how early childhood teachers use ongoing assessment for individualization of instruction. To identify studies for review, a library search was conducted that targeted research related to early childhood education (which we defined as including children from birth through 3rd grade) and early childhood special education. The search was limited to references from the previous 10 years (2002–2012). We also asked a team of experts who were consulting on this project to recommend seminal work prior to 2002 and articles that were in press in late 2012/early 2013. For a full list of search terms and parameters, see Akers et al. 2014.

The library search and expert recommendations identified 1,325 unduplicated references (1,281 references from the literature search and 44 from the expert recommendations). Three trained reviewers carefully screened all references for relevance. Based on criteria determined by the project team, this process resulted in 173 references screened as relevant for this review. Of the 173 studies in the review, nearly half (48 percent) were empirical studies (see Table I.1). The

[4] In Head Start, a child's first developmental screener is required to be conducted within 45 days of enrollment.

empirical studies included 56 descriptive studies (of which 25 were psychometric), 15 randomized controlled trials (RCTs), 7 quasi-experimental designs (QEDs), and 5 single-case designs (SCDs). Of all the studies in the review, 36 percent were conceptual pieces, 13 percent were guides that provided overviews of best practices or standards, and 2 percent were literature reviews or reviews of measures.

Table I.1. Designs of the studies identified by the literature review

| | Percentage (number) of studies | | | |
|---|---|---|---|---|
| | Total[a] | Early elementary | Preschool | Infant/toddler |
| Study design | | | | |
| Empirical | 48 (83) | 45 (41) | 46 (37) | 37 (13) |
| Descriptive—non-psychometric | 18 (31) | 16 (15) | 18 (14) | 11 (4) |
| Descriptive—psychometric | 14 (25) | 14 (13) | 16 (13) | 14 (5) |
| RCT | 9 (15) | 9 (8) | 6 (5) | 6 (2) |
| QED | 4 (7) | 3 (3) | 5 (4) | 0 (0) |
| SCD | 3 (5) | 2 (2) | 1 (1) | 6 (2) |
| Conceptual | 36 (63) | 41 (38) | 35 (28) | 31 (11) |
| Guide (best practices/standards) | 13 (23) | 12 (11) | 14 (11) | 26 (9) |
| Literature review or meta-analysis | 2 (4) | 2 (2) | 5 (4) | 6 (2) |
| Domain[b] | | | | |
| Language/literacy | 47 (81) | 46 (42) | 50 (40) | 37 (13) |
| Mathematics | 16 (28) | 23 (21) | 11 (9) | 3 (1) |
| Social and emotional | 16 (28) | 12 (11) | 15 (12) | 31 (11) |
| Science | 2 (3) | 2 (2) | 1 (1) | 0 (0) |
| Motor development | 1 (2) | 1 (1) | 1 (1) | 6 (2) |
| Not specified | 29 (51) | 25 (23) | 33 (26) | 29 (10) |
| Selected characteristics of the target population | | | | |
| Children who have disabilities | 40 (69) | 43 (40) | 33 (26) | 26 (9) |
| Children in Head Start or Early Head Start | 20 (34) | NA | 43 (34) | 34 (12) |
| **Total Number of studies** | **173** | **92** | **80** | **35** |

Note: Percentages may not total to 100 because of rounding.

[a] Thirty-four studies reported on more than one age group and are double-counted in the three age-specific columns; therefore, the total number of studies in the three age-specific columns sums to more than 173.

[b] Some studies did not report domain, and others reported on more than one domain.

NA = not applicable; RCT = randomized controlled trial; QED = quasi-experimental design; SCD = single-case design.

The distribution of study designs at both the early elementary and preschool levels approximately mirrors the distribution across all 173 studies, with empirical and conceptual studies together making up more than 80 percent. Of studies at the infant and toddler level, fewer were empirical, and more were guides than at the other two age levels (about one-third were empirical, approximately one-third were conceptual, and about one-quarter were guides that presented best practices or standards for implementing ongoing assessment). Across and within all age groups, studies most commonly discussed the use of ongoing child assessment in the

domains of language, literacy, or reading (47 percent of all studies). Note that 29 percent of the studies did not specify a domain and that some addressed more than one domain.

Overall, limited rigorous evidence is available about the areas critical for the successful implementation of ongoing child assessment to individualize instruction. However, the existing literature provides some recommendations for how ongoing assessment should be used for individualization and also describes the range of activities we are likely to see in early childhood settings:

- **Although the empirical literature on selecting an assessment method is limited, there are recommendations for the preferred characteristics of ongoing assessment.** Studies recommend that assessment methods be authentic (that is, collected in naturalistic contexts in everyday activities), ongoing, developmentally appropriate, individualized, and multifaceted (Bagnato et al. 2010, 2011; Pretti-Frontczak and colleagues 2011). Several studies recommend selecting assessment targets (that is, learning objectives to assess) that align with and measure critical outcomes of the curriculum, are teachable, are observable or measurable, and are generalizable (Hojnoski and Missall 2007; Good and Kaminski 1996; Good et al. 2001; Fuchs and Deno 1991; Bagnato et al. 2011; Hosp and Ardoin 2008).
- **Researchers recommend that teachers use multiple approaches to documenting and organizing information to support interpretation of child progress.** Methods for documenting information include checklists, ratings, anecdotal records, questionnaires, videos, and/or developmental scales. Examples of systems for organizing information include portfolios for compiling data from different sources; graphs; Excel spreadsheets, Access databases, or paper-based systems; and web-based or technology-enhanced systems to support documenting and organizing data.
- **Teachers use a variety of supports to interpret what is learned from child data and apply it to instruction.** To help them interpret data, teachers may rely on coaches or mentors, decision points set by programs, and web-based or technology-enhanced systems. When teachers use ongoing assessment data for individualization, they may use it to help them form and instruct small groups, create and implement tiered tasks or lesson plans, and identify children who need one-on-one assistance.
- **Families may be important partners in the collection and interpretation of ongoing assessment data.** Although the field lacks clear recommendations from recent empirical work, studies that discuss engaging families in ongoing child assessment describe families as important partners in the collection and interpretation of data.
- **Teachers may need support to overcome barriers to use ongoing assessment for individualization.** The literature suggests that (1) teachers may recognize the value of ongoing assessment, although they do not consistently collect assessment data, nor do they use it for instruction and individualization; (2) barriers to using data include lack of pedagogical content knowledge and lack of knowledge of assessment and interpretation of data; and (3) teachers want more training and professional development on the use of ongoing assessment to individualize instruction. Approaches to supporting teachers in making instructional decisions based on data include coaching and providing teachers with technology-enhanced systems to help interpret and use data. These systems may offer more immediate and tailored feedback to teachers and, in turn, can lead to better instructional

decision making and more positive outcomes for children. Comprehensive professional development appears to be more effective when it includes technology-driven support that provides immediate feedback (Al Otaiba et al. 2011; Buzhardt et al. 2011; Ysseldyke and Bolt 2007; Landry et al. 2009).

In sum, current research on ongoing assessment to individualize instruction is limited, yet provides important insights that help inform a conceptual model that focuses on understanding the use of ongoing assessment for individualized instruction and potential approaches to measuring those constructs.

**Box 1. Approaches to ongoing assessment in preschool**

| General Outcome Measures (GOMs) | Curriculum-embedded approach |
|---|---|
| An approach to formative assessment that produces data that describe the rate of growth across developmental or chronological periods. This type of measurement uses standard tasks that help teachers evaluate how well the instruction or intervention is increasing the probability of desired long-term outcomes. | An approach to formative assessment that draws on classroom activities as evidence of children's performance and progress. Varied methods are used to collect data, including anecdotal notes, photographs, samples of work, videos, ratings, and checklists. The data are compared to behaviors described in rubrics that identify whether a child is below, at, or beyond expected level of performance for age. |

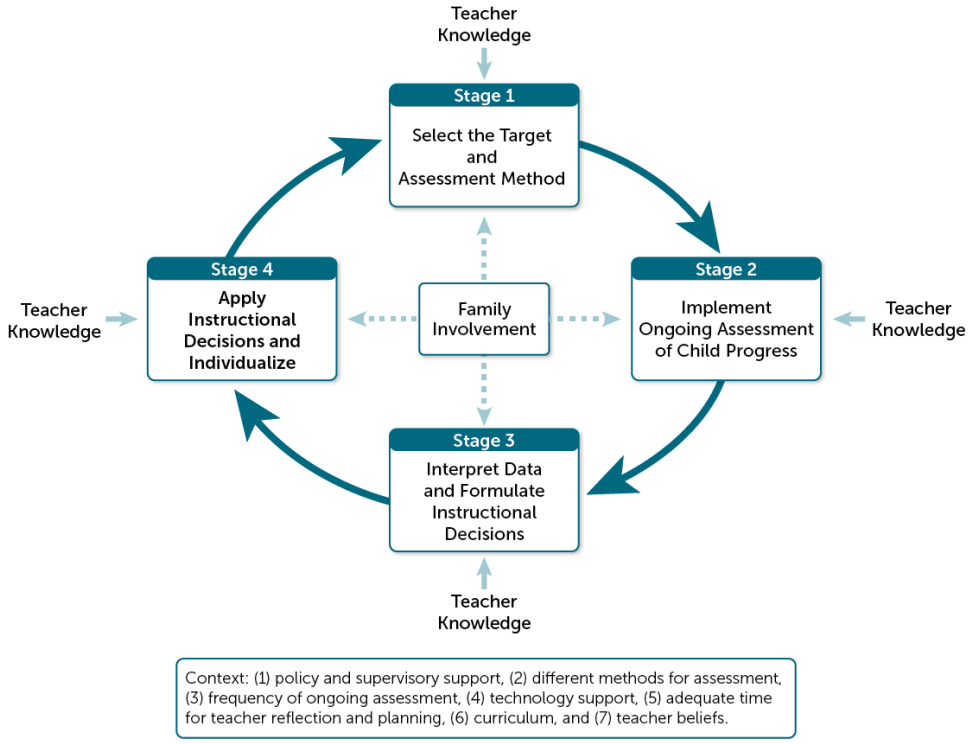
## C. Conceptual model that informs the EDIT

Evidence and theory identified through the literature review informed the project's conceptual model for the use of ongoing child assessment to individualize instruction, which, in turn, served as the foundation of a plan and a measure of teachers' implementation of ongoing assessment to individualize instruction and better meet children's needs. General Outcome Measures (GOMs) and curriculum-embedded approaches are two common approaches to ongoing assessment used in preschool classrooms (Box 1). The EDIT conceptual model focuses on a measure development plan for curriculum-embedded approaches because they are (1) more common in early childhood settings than GOMs; (2) more demanding for a teacher to implement (that is, they require greater teacher skills and knowledge); and (3) more comprehensive, because they traditionally cover several domains of development.

The conceptual model for using curriculum-embedded approaches to monitor children's progress on an ongoing basis and individualize instruction has four iterative stages (Figure I.1). Stage 1 is selecting the assessment target and method. Stages 2 and 3 are implementing the assessment and interpreting the assessment data, including hypothesis setting and selection of instructional decisions. Stage 4 is applying instructional decisions, including the collection of data to evaluate the success of the instruction. The next sections describe each stage and its quality indicators (for more details, see Atkins-Burnett et al. 2014). In addition, these four iterative stages are affected by contextual factors such as the policy and supervisory support that influence the teachers' data collection and interpretation, the different methods of assessment used across various settings, the frequency of the ongoing assessment, technical support provided to teachers, and whether teachers have enough time to collect data and reflect and plan instructional modifications. The type of curriculum being implemented, and teachers' beliefs about the assessment tools and about how children learn and what strategies are most effective,

also affect the collection of data and implementation of instruction based on ongoing assessment findings.

Figure I.1. Conceptual model for curriculum-embedded approaches




## 1. Stage 1: Selecting the target and assessment method

Program staff, rather than teachers, usually select the assessment system. However, teachers have some autonomy in selecting the assessment target (the skill, knowledge, or behavior to be assessed on a given day) and the assessment method (how that skill, knowledge, or behavior will be assessed), although both are also influenced by the assessment system and often by the curriculum. The selected assessment targets should represent the skills or knowledge a child would need to meet specified end-of-year goals.

There are several indicators of quality to consider when examining a teacher's selection of assessment target and method. These include whether the targets are meaningful, observable, responsive to instruction, able to show change during the preschool year, and able to be generalized beyond a specific context. The data collection method should be valid and collect information about the learning objective or behavior in reliable and efficient ways.

## 2. Stage 2: Implementing ongoing assessment

With assessment data collected during instructional activities, these assessments should have ecological validity, which means they should mirror what would actually happen in a real-world classroom setting. Teachers should implement efficient ways to collect data so that they can maximize instructional time. Teachers also need to document what children do objectively, accurately, and with relevant contextual information.

Indicators of quality to be measured at this stage include whether the teacher's implementation of the assessment is valid, replicable, individually appropriate, and fair, as well as whether the documentation is objective, complete, efficient, and consistent.

### 3. Stage 3: Interpreting data and formulating instructional decisions

Teachers need to be able to interpret the data about each child's performance relative to expectations for performance, usually based on data from typical same-age peers or developmental or curricular guidelines. Similarly, teachers need to be able to interpret each child's progress relative to developmental expectations. Each child's data, combined with other available data (such as information on instructional activities, peers' performance, national benchmarks, and family input), would help the teacher identify the child's strengths, weaknesses, interests, and learning differences and then select the best way to support the child's continued progress. The process of interpreting data and making instructional decisions may be conducted in teams with the support of other teachers, coaches, consultants, and family members.

Indicators of quality to be measured at this stage include how well the teacher has organized the assessment data, as well as whether interpretations are evidence-based and consider context and alternative hypotheses. Using data-based interpretations, the teacher then makes instructional decisions that are responsive to the data and draw on evidence-based instructional strategies as much as possible. After teachers try the evidence-based instructional strategy, they evaluate the effectiveness of the strategy for the child or children. There is a continuous feedback loop of evaluation, individualizing instruction, and assessing the effectiveness of the instruction to promote successful learning in children. Instructional decision making also involves planning when and what to collect to determine whether the instructional approach is beneficial to the child or children.

### 4. Stage 4: Applying instructional decisions and individualizing

When applying instructional decisions, teachers implement the instructional strategies with fidelity, the content is correct, and the level of rigor meets each child's needs. Individualization involves tailoring the instruction for individual children, although the instruction often is delivered within small groups. The teacher collects further evidence to evaluate whether the instructional approach is valid for meeting the targeted instructional need for each child. The teacher may group children with similar needs and may differentiate instruction in response to a child's needs. The teacher notes instructional changes and assesses progress to evaluate the success of those changes. The classroom instructional team members share their knowledge about each child's goals, instructional strategies, and progress.

The indicators of quality to consider at this stage include whether the teacher uses evidence-based strategies, differentiates instruction using a variety of approaches, uses instructional strategies that build on children's strengths and interests, and collects assessment data to evaluate the success of the instructional strategies.

## D. Contextual, teacher, and family factors affecting most stages

Curriculum-embedded approaches usually require teachers to make their own decisions about data collection, documentation, interpretation, and application. This means that the knowledge, opinions, and beliefs of teachers (and any other decision makers) strongly influence

the overall process. The context in which assessment occurs also affects the quality of implementation, so it is also important to consider key contextual factors that will help or hinder ongoing assessment. Accordingly, the curriculum-embedded conceptual model identifies factors with implications for the entire process: (1) teachers' knowledge and beliefs about assessment, instruction, and children's development; (2) family involvement in the process of ongoing assessment; and (3) available supports.

Teachers who believe that using data leads to better outcomes for children are more likely to collect and use data (Gallagher et al. 2008; Martin 2012; Sikka et al. 2007). Teachers with a solid grasp of pedagogy (how to teach) and child development are more skilled in all stages of the process, from selecting important and valid assessment targets aligned with the curriculum to individualizing instruction to meet children's needs (Buysse et al. 2013). Such teachers also will have the ability to share results with families and engage them in the process.

Families may become involved at several points in the process (although usually not when the data are being organized). At a minimum, teachers should keep family members posted on the child's progress and collaborate with them to interpret data when the child is struggling.

Policy and supervisory support for conducting *frequent* ongoing assessments is a key factor for implementation. This includes the availability of training, coaching, technology (such as access to smart phones or computers), or even supports as specific as clipboards and materials for organizing children's work. Quality implementation requires enough time for reflection and planning, as well as access to information on instructional strategies that are evidence-based or professionally recommended and aligned with the curriculum.

## E. Measurement model for the EDIT

The literature provided no evidence of a specific measure that could be used to assess teachers' use of ongoing assessment to inform instruction across different types of assessments. There was some guidance for how to evaluate teachers' use of assessments to inform instruction, but most of the methods were specific to a single assessment. Several studies measured fidelity of implementation for a specific assessment, usually a standardized assessment tool or a web-based system (Greenwood et al. 2011; Bolt et al. 2010; DeBaryshe et al. 2009; Landry et al. 2009; Grisham-Brown et al. 2008; Carter and Horner 2007; Ysseldyke and Bolt 2007; Fuchs et al. 1991; Hagans 2008; VanDerHeyden et al. 2008). Most of these studies used a checklist or a count of teacher behaviors that were specific to the assessment system. Six studies examined teacher reliability in scoring, but none of these used a curriculum-embedded assessment approach (Greenwood et al. 2011; Lo et al. 2009; Luze and Hughes 2008; Fuchs et al. 1991; VanDerHeyden et al. 2004, 2008). Teacher interviews using semi-structured protocols and/or think-aloud protocols were used in some studies (for example, Goertz et al 2009; Roehrig et al. 2008). One study also used a test of teacher knowledge—specifically, pedagogical content knowledge and assessment (Goertz et al. 2009). That same study used a multi-method approach, including conducting frequent observations of both instruction and assessments, and teacher interviews.

The resulting measurement model (Atkins-Burnett et al. 2014) elaborated on the constructs identified in the conceptual model and drew on the methods identified in the research studies, supplementing as needed with additional methods. The measurement model called for a

combination of checklists, ratings, and behaviorally anchored rubrics. Recommended data sources included review of planning, assessment, and instructional documentation; observations of teachers' instruction and implementation of ongoing assessment; and teacher interviews. Chapter II of this report describes the procedures used in the EDIT.

## F. Road map for the report

This chapter provided the conceptual basis for development of the EDIT. Chapter II describes the rationale for three data sources, summarizes the measure, and presents the administration procedures. Chapter III details the three rounds of data collection and the iterative process used to develop the current version of the EDIT. Chapter IV describes the pretest sample and what we learned from testing the EDIT. Chapter V discusses how these lessons affected the iterative development of the items and procedures of the EDIT. In Chapter VI, we propose possibilities for future work on the measure.

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

## II. THE EXAMINING DATA INFORMING TEACHING (EDIT) MEASURE: INSTRUMENT AND ADMINISTRATION

The EDIT tool is a multi-method measure of preschool teachers' use of curriculum-embedded approaches to ongoing child assessment and individualized instruction. The EDIT consists of a document review, video-based observations, and a one-hour teacher interview with reflective think-aloud probes.[5] Development of the EDIT draws on information from the literature review, input from an expert consultant group and ACF, and reviews of manuals for curriculum-embedded assessments. Together, this information helped the team identify key constructs to measure, as well as data sources for measuring them. The EDIT is grounded in the quality constructs identified in the conceptual model for the curriculum-embedded approach. Development of the EDIT balanced the competing considerations of (1) reliability and validity; (2) burden on teachers, classrooms, and raters; and (3) feasibility concerns, logistically and in terms of cost for both development and ongoing use. Overall, the project's goal was to develop the EDIT such that it could be used easily by researchers in diverse settings, to yield high quality data.

This chapter first describes the rationale behind the EDIT's multi-method approach and measurement issues considered during development, as well as its focus on two learning domains. We then provide an overview of the EDIT's three data sources and describe the number and type of items and their relation to the conceptual model. Finally, we present administration procedures for the EDIT.

### A. Rationale for a multi-method measure focused on two domains

In this section, we first discuss the need for a multi-method approach and then provide an overview of the EDIT's three data sources.[6]

#### 1. The need for a multi-method approach

Experts in research design and methodology strongly recommend using a multi-method approach when measuring constructs (Brewer and Hunter 2006). Specifically, many experts recommend triangulated measurement (Campbell and Fiske 1959; Denzin 1978; Denzin and Lincoln 2011; Patton 2002; Ritchie and Spencer 2002; Webb et al. 1966), which attempts to pinpoint a construct more accurately by approaching it from different methodological perspectives. To be useful and valid, an assessment must both provide consistent results and measure the phenomenon it intends to measure. When the different methods yield similar results, there is more confidence that the construct is being measured with validity.

[5] Throughout the semi-structured interview protocol, the rater uses probes asking teachers to reflect and "think aloud" about the documentation and videos, including how they made decisions as they conducted assessments and used the data to inform their teaching. Throughout this report, the term "think-aloud" refers to this reflective process.

[6] The EDIT measures how well a teacher conducts ongoing assessment for individualization, regardless of the assessment tool used. In other words, the EDIT is intended to rate the quality of a teacher's use of the tool, rather than the quality of the tool itself. However, some assessment tools might facilitate higher scores on particular items. For example, a computerized assessment tool may automatically organize data for interpretation.

Given the complexity of teachers' use of ongoing assessment for individualization, a multi-method approach best enables the EDIT to cover all aspects of the process and yield more complete information. This potentially results in a richer sample of what teachers think, know, and do when assessing children and tailoring instruction. The documents, for example, provide data on how teachers plan for instruction. They also shed light on teachers' understanding of how the content of activities and their structure (such as the use of small groups) can be used to tailor instruction. However, documents alone, without teacher feedback, do not provide the rich information needed to understand how the documents are being used to inform instruction. Similarly, the video recorded observations provide insight into teachers' actual instructional practice. However, incorrect inferences could be drawn from the videos if the EDIT did not include teacher reflection and input about the observations. The teacher interview adds this critical insight. Ultimately, each source contributes unique data and overlapping information that, together, describe how teachers collect ongoing assessment data and use that data to individualize instruction.

Several factors can affect the processes involved in teachers' use of ongoing assessment data, and it is important to recognize the different data sources where this use of assessment data may be occurring:

- **Teacher knowledge (interview and, to a lesser extent, the document review, and observation).**[7] Understanding what teachers are thinking as they complete assessments and individualize instruction sheds light on how teachers use their knowledge of child development, effective practices, individualization strategies, and the curriculum to make instructional decisions, as well as how they use child data to inform instruction and individualization.
- **Teacher practice (document review, observation, and, to a lesser extent, the interview).** Observing what teachers actually do in the classroom helps us understand how they apply what they know and believe, including how they implement curricula in general and how they collect and use ongoing assessment data more specifically. Teachers may be able to say what to do, but not know how to implement the processes. Teachers need to know how to translate their knowledge of children and assessment and instructional strategies into actual practice that supports the development of children. Teachers need to be able to adapt instruction when children are still not meeting performance or progress expectations. Classroom observations are the most direct way to measure this. In combination with teacher interview questions and child documentation, the classroom observations provide insight into additional factors that can indirectly affect the assessment and individualization process, such as classroom management, available classroom resources (such as an engaged assistant), and opportunities and challenges related to incorporating assessment and individualization into the classroom schedule.
- **Contextual factors (teacher interview and questionnaire).** To interpret the results, the team collects information on some contextual factors. For example, if a teacher is not modifying instruction based on the ongoing assessment data, this could be due to lack of knowledge, teacher beliefs about child development, or difficulty understanding how to

[7] Teacher beliefs can also influence teachers' use of ongoing assessment data. However, the EDIT focuses on teacher knowledge and practice because they are most likely related to effects of this process on children.

apply knowledge. Collecting information on how much and what type of professional development the teacher has received about ongoing assessment will help in understanding whether the difficulty is limited knowledge about assessment. The EDIT gathers some information about this through the teacher interview. During the pretest, however, the team relied primarily on a self-administered questionnaire (SAQ) to collect information on context, including the teacher's education and experience, as well as availability of mentoring, coaching, or other supports in assessment and instruction (Appendix A). The EDIT, combined with other data sources such as the SAQ, can help researchers and others understand how teachers' knowledge, beliefs, and ability to implement practices interact with the context to support or hinder ongoing assessment and individualization. This information could allow programs to provide targeted support and training to each teacher. However, the benefits of having this contextual information must be weighed against the additional time burden that collecting it places on teachers.

## 2. Measurement issues during the development phase

When developing a measure, it is necessary to balance the most valid and reliable approach with something that is feasible to implement (both logistically and in terms of cost), least intrusive, and most likely to provide critical information. Next, we highlight major measurement issues considered when designing the EDIT.[8]

**Balancing validity, reliability, and feasibility.** To ensure that the EDIT is valid and reliable, creation of the measure needed to include (1) multiple methods; (2) items or indicators that adequately represent the constructs being measured; (3) an examination across multiple learning domains to assess whether practices generalize across domains; and (4) adequate guidance for scoring (for example, detailed scoring rubrics for each construct) to facilitate rater reliability. However, we needed to balance these requirements with the need for feasibility, especially with the complexity of the multi-method approach and the amount of time it might take to train raters to learn and implement the different components.

**Burden.** The team used a multi-method approach that tried to minimize burden on teachers and raters. The team used several strategies to minimize the time and effort required of the teacher: (1) limiting the request for documentation to a specified time frame and requesting documents that the teacher already had; (2) allowing a two-week period for video recording observations, giving the teacher flexibility in his or her schedule; and (3) limiting the teacher interview to one hour, scheduled at the teacher's convenience. The team also limited the pretest to examining two learning domains: (1) language and literacy, and (2) social-emotional (discussed in more detail below). Using only two learning domains not only helped minimize the burden on the teacher, but also helped decrease the raters' cognitive burden and training burden. The team used rubrics, ratings, and checklists to balance the need to capture complex constructs with the cognitive burden placed on raters. Through the pretest, the team iteratively refined each data collection method to include only items that contributed unique and meaningful information.

[8] These considerations were specific to the measurement development phase. When the measure is finalized and implemented for research, the measurement considerations may differ.

**Accommodating a variety of assessment systems.** The measure needed to accommodate a variety of assessment systems used in Head Start classrooms, including electronic data collection and portfolios (such as Teaching Strategies: GOLD Online [Teaching Strategies, Inc. 2011] and Work Sampling System Online [Meisels et al. 2001]) and hard-copy records of child performance and behavior. Interview questions needed to be semi-structured to apply across different assessment systems, and raters bore a greater burden in additional training about when to ask follow-up questions and how to rate responses. One implication of developing the measure to accommodate a variety of systems is that raters will need a basic familiarity with a variety of ongoing assessment systems. The scoring system needed to allow for coding of both electronic and hard-copy documentation and account for the fact that teachers with electronic systems may make fewer decisions themselves (for example, the software may determine the organization of data and may indicate when a child fails to meet age-related performance expectations).

**Timing and frequency of observations.** The project team and the expert work group determined that the best way to gather information for the EDIT was to conduct at least three observations to help ensure that what was observed was a valid representation of teachers' practices. However, the timing of these observations affected the burden on teachers and their involvement in the project. Multiple in-person observations would have been costly and would not have been realistic for a measure brought to scale; therefore, the team opted to have teachers video record their assessment and small-group instructional activities multiple times over the course of two weeks. Allowing the teacher to video record also gave him or her more control over the timing of the activities and the selection of footage submitted to the rater.[9] Video recording also increased the feasibility of the multi-method measure because it required only a single visit by the rater to view the footage.

**Child selection.** One consideration was whether the classroom observation and document review should include the whole class or only selected children. We asked each teacher to use her judgment in selecting two children. Purposively sampling two children was more cost-effective and allowed for more in-depth study than examining the entire classroom. Including more children would have limited the EDIT raters' ability to carefully review data for each child. Selecting two children increased the raters' ability to examine more data for each child and helped the rater more easily understand the teacher's use of assessment data to determine child progress.

### 3. Initially targeting two domains

All the domains in the Head Start Child Outcomes Framework are important for understanding child progress and, therefore, are relevant to this study. However, several constraints prohibited exploring all domains in the initial phase of measurement development. Perhaps most prominent is the burden on teachers, classrooms, and raters when data are collected in many domains.

[9] Allowing teachers to select their own footage may introduce bias (because teachers may submit what they perceive to be their best footage), and this bias may limit learning about actual teacher practices. However, the footage can still capture teachers' knowledge of assessment because, to select the best footage, the teachers must understand what constitutes good assessment practice.

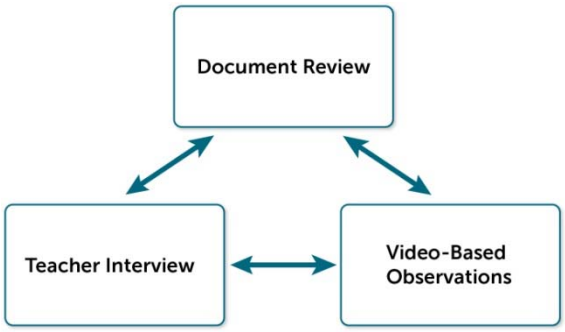
Instead of imposing this burden, the expert panel suggested pretesting only two domains: (1) language and literacy, and (2) social-emotional. Both are significantly linked to long-term well-being. For example, research shows that early competency in language and literacy is tied to later success in school (Lee and Donahue 2007; Rowe et al. 2012). In addition, Head Start has invested considerable resources in training teachers to teach language and literacy. Given this investment, it is important to know whether teachers can use what they have learned in their trainings to individualize instruction. The other domain—social and emotional—often has an overarching impact on children's behaviors and cognitive processing (Heckman and Raut 2013), because a child with social and emotional problems may also have difficulty learning. In addition, early childhood teachers are more likely to deliver instruction in language and literacy than in other areas such as mathematics, and they typically stress literacy and social and emotional development (National Research Council of the National Academies 2009). Therefore, the teachers were likely to provide enough instruction in the documentation and video-recordings to observe whether the teacher varies instruction for different children. Finally, limiting the number of domains allowed the team to thoroughly refine the measurement of those domains within the project's resources.

Despite limiting the focus to two domains, items are included in the EDIT to examine whether teachers collected documentation across five Head Start domains and drew on all domains to interpret child data and individualize instruction (for example, whether a teacher employs strategies that encourage persistence—part of the approaches-to-learning domain—during a literacy activity). In the future, the EDIT could be expanded to cover assessment of additional domains. This would help reinforce the message of a "whole child" approach and avoid giving the impression that language, literacy, and social and emotional development are the only important domains.

## B. Overview of the three methods

As Figure II.1 shows, the multi-method approach of the EDIT allows for triangulation across data sources. When teachers participate in the EDIT, they gather documents and create video recordings over a two-week period, after which the EDIT rater conducts a site visit[10] to review the documents, watch the video, and interview the teacher. Next, we discuss each of the EDIT methods in detail.

Figure II.1. Multi-method measure model




[10] Depending on the amount of data provided by the teachers in the documents and videos, the site visit during the development phase required between four and six hours in a single day.

### 1. Document review

Teachers provide documentation for the two focal children: one performing well in language and literacy and another facing challenges.[11] Examples of documentation include student portfolios, assessment records, and lesson plans (Exhibit II.1). The rater reviews the documents to examine the intentionality, focus, completeness, and objectivity of the teacher's data collection and instructional planning. This review can address constructs from each of the four stages in the conceptual model.

Exhibit II.1. Examples of documentation

**Assessment documents**

- Assessment schedules and plans
- Portfolios (samples of children's work, photos, and other teacher documentation)
- Assessment records (such as checklists and anecdotal records)
- Assessment reports

**Instructional documents**

- Lesson plans and plans for individualized instruction
- Curriculum/instructional sequence
- Goals/objectives for child learning and development

The document review focuses on two types of documents: (1) assessment documents (such as plans for assessments and assessment results), and (2) instructional documents (such as lesson plans). For assessment documents, the rater reviews what and how data are collected, documented, and organized by the teacher (for example, evidence of frequent data collection in an assessment plan, objective documentation in anecdotes, and data organization in data displays). For the instructional documents, the rater reviews teachers' lesson plans for evidence that their instructional plans and any associated plans for individualization are responsive to the documented data, and draw on evidence-based or professionally recommended strategies.

### 2. Video-based classroom observations

Over the two weeks before the raters' visit, the lead teacher in each classroom uses a tablet to video record a selection of approximately six activities that varied in time and involved at least one ongoing assessment and two small-group instruction activities with each of the focal children. Teachers collect assessment data as they normally do. For the small-group instruction, the team asks that language or literacy be the focus, but otherwise does not put parameters on the activities, instead asking teachers to follow their typical classroom practice.

[11] The team experimented with focusing on ratings of children's social-emotional development or language and literacy for selecting the focal children. Because teachers tended to have the most data points for language and literacy, we are moving forward with selecting on the basis of language and literacy. However, during the document review, video-based observations, and teacher interview, the raters are attentive to children's development in all domains, including the social-emotional domains. Only one teacher had an individualized goal for a child in the social-emotional domain.

Video recorded observations provide a critical piece of information about how teachers conduct assessments and use the data from the document review to inform their instructional practice. The rater examines how teachers collect data and individualize instruction. The rater views the recordings during the site visit after the document review and before the teacher interview. The videos can be used to provide information on stages 2 (implementing ongoing assessment) and 4 (applying instructional decisions and individualizing) of the conceptual model, which focus on teacher implementation and use of assessments.[12]

### 3. Teacher interview

The rater interviews the lead teacher at the end of the site visit. The interview includes a standard set of questions about the teacher's use of ongoing assessment data to inform instruction. For example, the rater asks about the teacher's decision making regarding the types of information to collect for each child, as well as how and when to collect and record that information. Other questions ask the teacher about how she organizes and interprets the data, such as how she decides whether a child is making enough progress on a given learning objective and whether a child needs additional support or a new instructional approach. The rater probes about how—and how often—the teacher examines the data (such as looking at the data by child and objective) to gauge child performance and progress. The rater also asks about how the teacher uses assessment information to inform instructional decision making—including lesson planning and the use of adaptations, modifications, and individualized teaching strategies—and to gauge the success of instructional interventions.

In addition to the standard questions, the rater embeds reflective think-aloud probes throughout the interview to better understand the teacher's thinking and decision making around the documents and video recorded observations. Before the interview starts, the rater organizes these questions to align with stages in the conceptual framework. These questions are non-standard probes that will vary based on the documents and video contents. For example, reflective think-aloud probes may ask the teacher to describe how she uses the documentation that she collected, as well as her thoughts and actions as she planned and implemented the video recorded tasks. The rater may ask questions about:

- What the documents obtained as part of the document review reveal about a child's abilities and any steps the teacher took or planned to take to support the child's learning based on that understanding. The rater may use additional probes, but needs to take care to avoid leading teachers and skewing the results.
- Why the teacher collected the particular information that she did (listening for intentionality and how it relates to the curriculum and to each child's needs); how she interpreted the data; and what actions, if any, she took in response to what she observed (specifically listening for examples of individualizing instruction).
- Whether the teacher shared information with parents or involved them in collecting and interpreting any data. If the teacher does share the data or otherwise involve families, the rater asks for specific examples of how the teacher does this.

[12] To augment what the rater sees in the video, the interview includes reflective think-aloud probes to provide more insight on the teacher's thoughts and actions.

- Whether similar information was collected for other children in the classroom and how the teacher used information from different children to make decisions (for example, to group children or to plan instruction).

The rater uses the teacher's responses to the think-aloud probes and interview questions to finalize scores on all items. Interview responses address constructs in all four stages of the conceptual model.

## C. Summary of the measure

In this section, we first describe the number and types of items and then present the items' relation to the conceptual model.

### 1. Number and types of items

The EDIT instrument consists of five holistic rubrics, two sets of analytic rubrics, three sets of ratings, and two checklists, for a total of 54 items (Exhibit II.2 and Table II.1).

Exhibit II.2. Definitions of EDIT item types

| Definitions of EDIT item types | |
|---|---|
| Rubrics | *Rubrics* are scoring guides that describe several levels of quality and multiple aspects of performance. They can be developed to document fine gradients of change to capture small but important differences across various aspects of quality. Rubrics incorporate ratings (for example, a scale of 1 to 5) but are more elaborate than ratings, often including multiple criteria that must be met before an item can earn points on the rating scale.<br>*Analytic rubrics* are used to rate individual dimensions of quality (for example, to rate the flexible use of instructional strategies separately from reflection on the success of the strategies).<br>*Holistic rubrics* are used to examine multiple dimensions or characteristics that co-occur (for example, to rate the flexible use of instructional strategies along with reflection on the success of the strategies and other aspects of individualizing instruction). Rubrics can be used for multidimensional concepts, such as examining how instruction is individualized to accommodate each child's strengths and weaknesses.<br>Rubrics are useful for rating qualitative differences and can provide especially rich representations of a teacher's practices. Analytic rubrics are best for providing feedback to teachers and informing professional development, whereas holistic rubrics are particularly useful when the whole is more than the sum of its parts. However, the subjective nature of rubrics necessitates more training to obtain inter-rater reliability compared to checklists or ratings. Rubrics may also combine quantitative and qualitative criteria, sometimes making it difficult to weigh multiple dimensions of quality within the same observation (for example, if the teacher individualizes appropriately with some children but not with others). Rubrics are also more time-consuming for the rater to complete than the other methods, which may be burdensome. However, qualitative rubrics of teacher practice are usually more strongly associated with child outcomes than other types of measurement, such as checklists (Chomat-Mooney et al. 2008). |
| Ratings | *Ratings* are scales that take measurements along a continuum (for example, 1 to 5 or "not at all" to "always"). Rating scales can vary in length based on the desired number of gradations, and the descriptions of points along the scale should be clear and hierarchical. Rating scales may measure frequency (for example, how frequently a teacher observed a child's skill in a particular domain). They could also measure how characteristic a behavior is (for example, "How characteristic of this teacher is the following statement: 'All of this teacher's documentation is objective'?" with ratings from "not at all" to "very characteristic"). Rating scales provide more detailed information than checklists. However, the subjective nature of assigning ratings necessitates more training to obtain inter-rater reliability, compared to checklists. |

| Definitions of EDIT item types | |
|---|---|
| Checklists | Raters use *checklists* to identify the presence or absence of behaviors, skills, or documents. Checklists should include clear definitions of each item being checked. For example, the types of approaches used for individualization—such as additional practice, grouping strategies, and peer supports—could be items on the list. When accompanied by definitions, these items can be reliable and relatively objective measures; however, they do not capture gradations and qualitative content. |

Note: For additional information on measurement options selected for the EDIT, see Atkins-Burnett et al. 2014. The three item types were useful and were maintained in the current EDIT.

Table II.1. Number and types of items, by conceptual model stage

| Conceptual model stage | Conceptual model activity | Type of items | Number of items | Data sources |
|---|---|---|---|---|
| Stages 1 (selecting the assessment target and method) and 2 (implementing ongoing assessment) | Selecting the assessment target | Holistic rubric | 1 | Documentation, Interview |
| | Selecting and implementing the assessment method | Analytic rubric, ratings | 14 | Documentation, Observations, Interview |
| | Documenting the information collected | Ratings | 9 | Documentation, Observations, Interview |
| Stage 3 (interpreting data and formulating instructional decisions) | Organizing the data | Holistic rubric, ratings | 18 | Documentation, Interview |
| | Interpreting the data | Holistic rubric, analytic rubric, checklist | 3 | Documentation, Interview |
| | Formulating instructional decisions | Holistic rubric, checklist | 8 | Documentation, Interview |
| Stage 4 (applying instructional decisions and individualizing) | Applying instructional decisions and individualizing | Holistic rubric | 1 | Documentation, Observations, Interview |

Where possible, the EDIT uses holistic rubrics to capture the multidimensional constructs because they are most appropriate for qualitative subject matter in which the whole is greater than the sum of its parts. For example, even if the teacher's interpretation of the data is based on multiple pieces of information collected in more than one context or task, the interpretation will not be valid unless it also considers the context of observed skill/behavior. A holistic rubric can tease out these qualitative aspects of the process. The EDIT also uses analytic rubrics to rate individual dimensions of quality (for example, to rate the frequency of data collection separately from documentation of the context of the assessment). Checklists or ratings are used when the presence or frequency of certain behaviors is important (for example, how frequently the teacher enters assessment information into the assessment system). The EDIT captures the constructs in each stage of the conceptual model; the team used the level of measurement (e.g., checklist or rating) that best fit the construct within each stage, choosing the least burdensome level for raters when possible.

## D. Administration procedures

The EDIT is designed for preschool classrooms using a curriculum-embedded approach to ongoing child assessment. In this section, we discuss each step in the administration of the EDIT.

### 1. Selecting focal children

The team asks each teacher to select two focal children from consented children in the classroom.

### 2. Teacher activities before the raters' visit

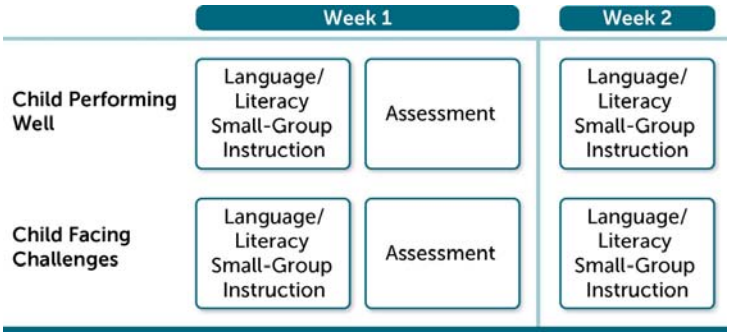
**Assembling documents.** Lead teachers in each classroom receive a letter explaining the study activities that includes a description of the documents that the team would like the teachers to collect before the raters' visit to the program (see Appendix B for the teacher letter). The letter specifies that the team would like to see any documentation that teachers collect and use to help them decide how to support the learning of the two focal children. Specifically, the team asks for information they collected on the two focal children from the two months before the site visit, as well as any plans for instruction (both classroom-level lesson plans and any individualized plans for those two children) from those two months. The team provides a list of example documents (such as anecdotal records, portfolios, family or class reports, lesson plans and individualized learning plans, instructional sequences, and assessment schedules). The team emphasizes that it wants to collect existing documents that are readily available and do not require additional teacher effort to produce. An additional benefit was that the raters noticed a greater difference between the children performing well and the children facing challenges when teachers selected focal children.

**Learning to use the video equipment.** Teachers are shipped an electronic tablet and accessories (including a charger and small flexible tripod) with written instructions for video recording. They are encouraged to experiment with recording before the formal recording of the video observation activities to familiarize themselves with the equipment and desensitize the children to its presence. In the letter, the team recommends that the teacher participate in a call with a member of the team to ask questions and receive technical assistance if needed.

**Collecting videos.** In each classroom, the lead teacher is asked to record a selection of activities involving individual ongoing assessment and small-group instruction with the focal children over the two weeks before the raters' site visit. The team asks that the teacher video record the following sequence (Figure II.2) for one child performing well and another child facing challenges, resulting in six data points collected across two weeks.[13]

[13] If the teacher would normally group the two focal children for small-group instruction, she may use the same small-group instruction videos for both children.

Figure II.2. Method for the focus and frequency of video recordings




For the small-group instruction, the team asks that language or literacy[14] be the focus but otherwise does not restrict the activities, instead asking teachers to follow their typical classroom practice. For the assessment activities, the team asks that teachers use the assessment tool typically used in their classroom.

### 3. EDIT rater(s)' visit

During the visit, the rater:

- Reviews documents to see how the teacher is using ongoing assessment data to individualize instruction
- Reviews the video recorded observations to see how the teacher administers assessments and individualizes instruction
- Conducts a one-hour individual teacher interview to probe for additional explanations about the documents and observations, as well as the teacher's planning and implementation of adaptations, modifications, and individualized teaching strategies

Table II.2 depicts an example of a schedule for the site visit at a part-day program.

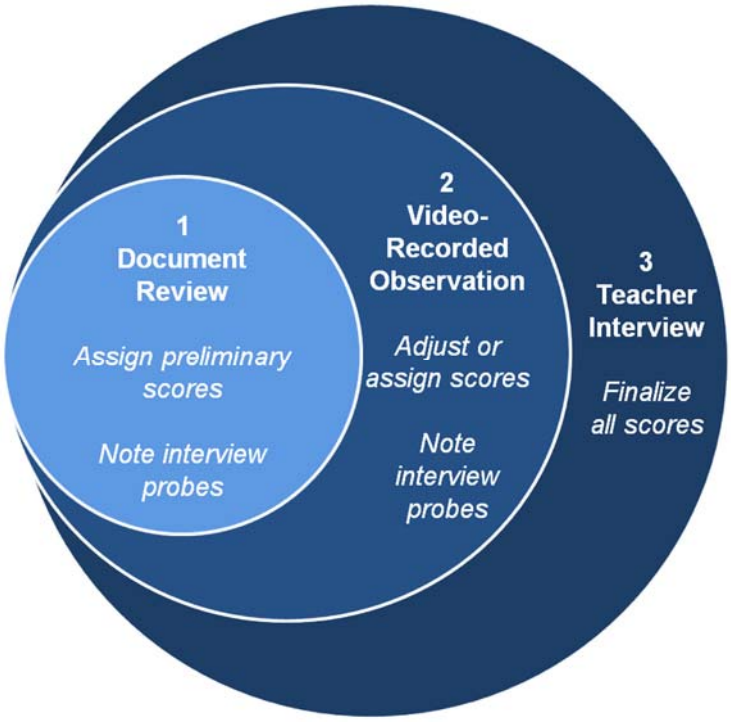
Table II.2. Proposed schedule for site visit for one part-day classroom

| Part-day program timing | Activity |
|---|---|
| 11:00 a.m.–1:30 p.m. | Rater reviews the documents and assigns preliminary scores where possible. |
| 1:30–2:30 p.m. | Rater reviews the video gathered by the teacher and adjusts or assigns scores as needed. |
| 2:30–3:30 p.m. | Rater prepares interview questions (to fill gaps or answer questions based on document review or video). |
| 3:30–4:30 p.m. (after program) | Teacher participates in a one-hour interview with the rater. |
| 4:30–5:30 p.m. | Rater finalizes all scores. |

[14] Language and literacy were selected because early childhood teachers are more likely to deliver instruction in language and literacy than in other areas (such as mathematics). Although we did not ask teachers to video record social-emotional instructional activities, raters did look for data related to social-emotional development in the documents. Only one teacher selected a child who had an individualized goal in the social-emotional area.

**Preliminary scoring based on data collected from the document review.** The rater completes the document review at the program during a scheduled visit. The document review requires approximately two and a half hours to complete. The rater notes the type and frequency of assessments, how data have been used to guide instruction, and any evidence of individualization in lesson plans. The EDIT allows the rater to rate whether the documentation for each child provides evidence of the child's unique strengths and weaknesses, as well as common areas of progress across children. The rater checks for information across time about specific areas of development, rather than on a random collection of skills and behaviors that do not focus on progress. Where possible, the rater assigns the teacher a preliminary score (Figure II.3). For example, item 1 (a holistic rubric for measuring 'selecting the assessment target') is initially rated based on the document review and that rating is revisited after the video and again after the interview. For items that cannot be scored without information from the video-based observations or the teacher interview, the rater waits to assign a preliminary score (for example, items 2f-2h are based only on the videos). The rater notes whether additional evidence is needed on any quality indicators; if so, the rater watches closely for that evidence during the video observations and/or probes for additional evidence during the teacher interview.

Figure II.3. Use of methods to assign scores




**Scoring data collected from the videos.** The EDIT rater views the recordings during the site visit (after the document review and before the teacher interview). The EDIT rater then adjusts scores on previously scored items as needed and assigns preliminary scores on remaining items. The EDIT rater notes whether to probe for additional information on any quality indicators during the teacher interview if these indicators have not been observed in the documents or videos. The EDIT rater collects the tablet at the end of the visit, and the videos remain available to the EDIT team for further discussion.

**Interviewing the teacher.** The EDIT rater conducts a one-hour individual interview with the lead teacher at the end of the visit to talk about the decisions made in collecting and using ongoing assessment data. The EDIT rater organizes the questions before the start of the teacher interview. To ensure that the interview can be conducted in person, the EDIT rater's visit is scheduled at the teacher's convenience, preferably on a day when an assistant can cover the time the teacher is participating in the interview. Conducting the interview during children's naptime sometimes is feasible. Conducting the interviews in the classroom is ideal, because the teacher has access to all classroom materials. The EDIT rater could also visit the classroom on a teacher in-service day or could conduct the interview during the evening, if that is most convenient for the teacher. Preferably, the interview is conducted in person so documents can be shared as needed. After the interview, the EDIT rater has an additional hour to finalize scores on all items.

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

## III. THE MEASUREMENT DEVELOPMENT PROCESS

We pretested the EDIT to examine how the items performed, refining the items as necessary throughout, and to examine how our proposed procedures worked. In this chapter, we outline pretest activities, beginning with recruitment. We then describe the pretest timeline and procedures, including activities that teachers completed in advance, as well as activities during and after raters' visits. We conclude by describing the process we used to refine the EDIT and its procedures.

### A. Recruitment

The iterative pretest consisted of three rounds of data collection conducted with nine English-speaking early childhood teachers in six Head Start Centers with video recorded observations of the teachers working with the children in their classrooms[15] (Table III.1).

Table III.1. Rounds of data collection for the pretest

| Center | Classroom |
|---|---|
| **Data collection round 1: March–April 2014** | |
| Center A | Classroom 1 |
| | Classroom 2 |
| Center B | Classroom 3 |
| | Classroom 4 |
| **Data collection round 2: May–June 2014** | |
| Center C | Classroom 5 |
| | Classroom 6 |
| Center D | Classroom 7 |
| Center E | Classroom 8 |
| **Data collection round 3: November 2014** | |
| Center F | Classroom 9 |

The EDIT development team sought centers using a curriculum-embedded ongoing assessment system with moderate to high quality of implementation. The team chose the centers purposively based on recommendations from research colleagues who had been in classrooms at the centers. In general, the researchers thought the centers we visited were implementing ongoing assessment with moderate to high fidelity. The centers were in three states.

[15] Two of the lead teachers were Spanish-English bilingual, and four of the classrooms had children from households with Spanish speakers.

## B. Pretest timeline and data collection procedures

Each round of data collection included pre-visit activities, a site visit, and post-visit debrief calls with teachers (usually two teachers) from that round. The length of each round of data collection varied, because the EDIT development team refined EDIT data collection procedures between rounds. In the sections that follow, we discuss specific activities related to recruitment, the selection of focal children, teacher collection of data before the site visit, the EDIT raters' visit, and debriefing with teachers. Table III.2 illustrates the differences between activities in each round in more detail.

Table III.2. Depiction of differences in pretest data collection, by round

| Data collection activities | Round 1 | Round 2 | Round 3 |
|---|---|---|---|
| EDIT advance letter to setting contact. | Δ | Δ | Δ |
| EDIT team calls setting contact and identifies SPP.* | Δ | Δ | Δ |
| If SPP is different from initial setting contact, EDIT team calls SPP. | Δ | Δ | Δ |
| EDIT team sends consent packets to SPP, and SPP distributes consent packets to teachers. | Δ | | Δ |
| Teachers/SPP collect consents and ship them to EDIT team. | Δ | | Δ |
| EDIT team sends OL&C** packets to teachers for consented children. | Δ | | |
| Teachers complete OL&Cs and send results to EDIT team. | Δ | | |
| EDIT team randomly selects two focal children stratified by performance on OL&C. | Δ | | |
| EDIT team ships teachers video recording materials with identification of focal children and schedule for video recording. | Δ | | |
| EDIT team ships teachers video recording materials and schedule with instructions to select two focal children and return consent for those children before beginning additional activities. | | Δ | |
| EDIT team ships teachers video recording packet with identification of consented children eligible to be selected as focal children. Teachers select focal children, collect documents, and record videos on assigned schedule. | | | Δ |
| EDIT team offers telephone support for teachers. | Δ | Δ | |
| EDIT team requests call with teacher to discuss video recording. | | | Δ |
| Teachers conduct video recording. | Δ | Δ | Δ |
| Teachers gather documents. | Δ | Δ | Δ |
| EDIT team conducts site visit with two raters. | ▲ | ▲ | ▲ |
| EDIT team holds debrief calls with teachers. | ☐ | ☐ | ☐ |

*Setting Point Person (SPP) is a primary point of contact at each center.
**Oral Language & Comprehension Questionnaire (adapted from Bradfield and McConnell 2013).
Δ Before visit ▲ During visit ☐ Post-visit

### 1. Recruitment and consent

The EDIT team began all recruitment efforts by sending an advance letter to each setting. A member of the team then called the setting contact (usually the program director or other

administrative staff) to discuss the study, verify the center's eligibility, and identify a Setting Point Person (SPP). The setting contact also identified the teachers to participate in the EDIT, and criteria for selection within each center varied, as it was at the discretion of the setting contact. Next, the EDIT team shipped consent packets—including Institutional Review Board (IRB)-approved consent forms[16]—to the SPP to distribute to the participating lead teachers, who, in turn, distributed consent packets to parents of the children in their classrooms.[17] The teachers returned consent packets to the project team. The team also collected consent forms for the lead teacher in each classroom selected to participate by the SPP. As Table III.3 shows, the targeted number of children with consent varied in each round of the pretest as different approaches on the schedule and data collection procedures were tested. In the first round of data collection, the team sought consent from all children in participating classrooms. Consent collection took several weeks. To test procedures, the team sought consent only from teacher pre-selected focal children in participating classrooms in round 2 of data collection. This saved time but limited the types of video recordings that could be collected. Teachers avoided small group instruction and sometimes video recorded outside the classroom because it was difficult to ensure that only consented children were visible in the video. In round 3 of the pretest, the team once again sought consent from all children in the classroom to allow teachers to video record more naturally occurring data collection and instruction even though it took more time to collect consent.

Table III.3. Consent collection during pretest

| Data collection round | Consent lead teacher | Consent all children in classroom | Consent only focal children |
|---|---|---|---|
| Data collection round 1: March–April 2014 | X | X | |
| Data collection round 2: May–June 2014 | X | | X |
| Data collection round 3: November 2014 | X | X | |

## 2. Selecting focal children

The team pretested a method for selecting two focal children (one performing well[18] and the other facing challenges in language and literacy) who would be video recorded and for whom the documents would be collected.

In the first round of data collection, we experimented with asking participating teachers to complete the Oral Language & Comprehension measure (OL&C; Bradfield and McConnell 2013) for consented children. This brief teacher-report measure has only eight items per child. It takes about 10 to 15 minutes to complete for an entire class and does not require training. It is specific to the domains of language and literacy and social-emotional development and behavior.

[16] Consent was sought from teachers and from parents for permission for the video-based observations and for the teacher to complete the OL&C measure for all children, and share documents about the focal children.

[17] Participating teachers were selected by the programs without input from the CPM study team.

[18] By "performing well," we refer to children meeting or exceeding developmental expectations for their age.

The EDIT team then used the results of this measure to rank order and stratify the children and randomly select one child performing well and one child facing challenges.[19] However, the OL&C was intended for a different purpose, specifically, identifying children who are struggling, but not necessarily ranking children who are performing well.

In the second and third rounds of data collection, teachers were not asked to complete the OL&C, and the EDIT team did not select the focal children. Instead, participating teachers were asked to identify one child performing well in language and literacy and another child facing challenges. The elimination of the mailing time both before and after completion of the OL&C reduced the time leading up to the site visit by one to two weeks.

### 3. Teacher activities before the raters' visit

**Assembling documents.** We sent a letter explaining the study activities to the lead teacher in each classroom. The letter included a description of the documents that we wanted teachers to assemble prior to the team's visit to the program. The letter asked teachers to gather any documentation that they had related to their assessments and plans for instruction for the focal children. The letter also included examples of the types of documents teachers might collect (such as assessment schedules, portfolios, anecdotal records, checklists, lesson plans, and assessment output/reports). The goal was for teachers to collect existing documents that were readily available and that did not require them to produce new documentation.

**Learning to use the video equipment.** Teachers also received a package with an electronic tablet and accessories (including a charger and small flexible tripod) and written instructions for video recording. They were encouraged to experiment with recording before the formal recording of the video observation activities to familiarize themselves with the equipment and desensitize the children to its presence. A member of the team was always available to talk with each teacher by telephone to answer questions or to offer technical assistance if needed. None of the teachers in the pretest elected to speak with a member of the team about using the video equipment.

**Collecting videos.** In each classroom, the lead teacher was asked to record a selection of activities involving ongoing assessment and small group instruction that included the focal children over a period of two to three weeks. Video recordings were completed before the EDIT raters visited the setting.[20] For the assessment activities, teachers were asked to use the ongoing assessment tool typically used in their classroom. For the small group instruction, we requested that language or literacy be the focus, but otherwise there were no parameters on the activities. Teachers were asked to follow their typical classroom practice. Teachers were able to include only consented children in the activities they video recorded.

During the iterative pretest, the team tried different approaches to video recording, including varying the content focus, timing, and number of videos per child. Each teacher was assigned to

[19] We also selected an alternate for each focal child in case one of the children could not participate in subsequent activities (for example, if the child was absent during the video recordings) or if a parent withdrew permission.

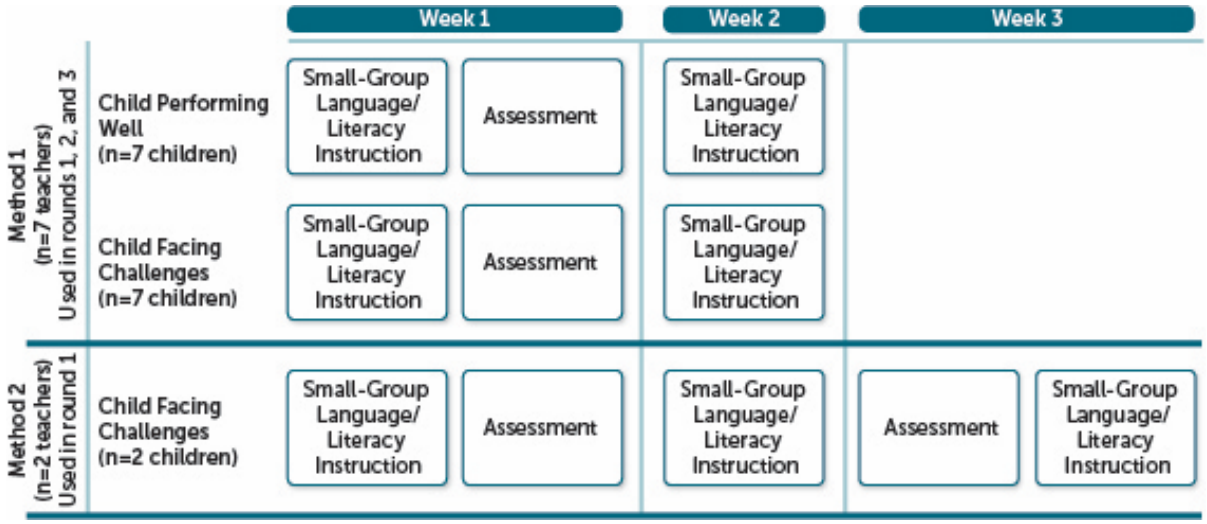
[20] The team also asked the teachers to check that the video recording was successful (that is, that the recording was not blank).

only one method. In the first round of data collection, the team examined two alternative video recording methods that varied in the instructional or assessment focus and frequency of the video recordings (Figure III.1):

- **Method 1:** The classroom teacher was asked to video record the following sequence for one high-performing child and one low-performing child, resulting in six data points collected across two weeks:
  - Week 1: Language and/or literacy small group instruction and an assessment
  - Week 2: Language and/or literacy small group instruction
- **Method 2:** The classroom teacher was asked to video record the following sequence for one low-performing child,[21] resulting in five data points collected across three weeks:
  - Week 1: Language and/or literacy small group instruction and an assessment
  - Week 2: Language and/or literacy small group instruction
  - Week 3: An assessment and language and/or literacy small group instruction

We opted to focus on a low-performing child in method 2 because the team assumed that changes made for low-performing children would be more obvious to the rater. In addition, it may be easier to see more progress in relation to the instructional changes across this brief time period with low-performing children than with children performing well. However, our raters noted that teachers often attend more to children who are having difficulty, so the results from method 2 might not generalize to typically developing children. Video recording methods were assigned by center in data collection round 1.

Figure III.1. Alternative methods for the focus and frequency of video recordings




[21] We reviewed the documentation for the high-performing child and included that child in the questioning during the interviews, but did not collect a video recording of that child in this method.

In rounds 2 and 3 of data collection, all teachers were asked to collect the same number and sequence of videos for each focal child (method 1 in Figure III.1). All teachers recorded a different number and sequence of videos than requested.

### 4. Raters' visit

During the pretest, two raters from the EDIT development team[22] visited each classroom for the site visits. On the site visit, raters reviewed documents collected by the teacher, reviewed video recordings, and conducted a one-hour interview with the lead teacher. Most members of the team had prior experience with early childhood classroom observation instruments. Although the raters scored the instruments independently, they conferred in planning questions for the teacher interview, and both were present during the teacher interview, debriefed after the classroom visit, and discussed scoring discrepancies in order to come to consensus on item-level scores.

### 5. Self-administered questionnaire

After the teacher interview, teachers were asked to complete a brief self-administered questionnaire (SAQ) to help the team learn more about teachers in the study (Appendix B). The SAQ included information on the teacher's demographics, early childhood background and experience, and classroom. Data on teachers are reported in Chapters IV and V.

### 6. Teacher debrief

After each round of data collection, all teachers were asked to participate in a 20-minute telephone call to debrief on the experience, and most agreed to do so. The debrief discussions gave teachers a chance to comment on the process, including burden for collecting consents, gathering documents, and ease of video recording.

## C. Refining the instrument and its procedures

After each round of data collection, the full EDIT team debriefed. During these discussions, the team considered changes to the items, protocols, and procedures. Recommended changes were presented to ACF, and feedback was incorporated. The process was repeated for each round of pretesting to iteratively develop and refine the EDIT. See Chapter V for a description of the revisions.

In between the second and third round of data collection, the EDIT development team shared the EDIT with the project's expert panel for their review. During a webinar with the expert panel, the team shared a summary of the pretest procedures; described the burden associated with each data collection method (for example, the average length of the teacher interview); and summarized the lessons learned from the first two rounds of pretest data collection with implications for the items and procedures. Feedback from the expert panel was incorporated for the final round of data collection.

[22] The EDIT development team consisted of six researchers. Five of the researchers paired in different combinations for each visit.

# IV. PRETEST

We begin this chapter by describing the pretest sample, including the characteristics of the teachers and classrooms that we visited. We then describe what we learned in those classrooms based on using the EDIT measure.

## A. Sample: Who participated in the pretest?

As noted in Chapter III, we conducted the pretest with nine classrooms from six Head Start centers, visiting eight classrooms in spring 2014 and one additional classroom in fall 2014. The spring visits took place between March and June, and the fall visit occurred two and a half months after the start of the year.[23] The classrooms were Head Start classrooms in three states, and some of the classrooms were part of grantees who blended funding and also were part of state-supported programs. Of the classrooms we visited, one was a classroom of 3-year-olds, four were classrooms of 4-year-olds, and four were classrooms of 3- and 4-year-olds. Four of the classrooms had children who were Spanish-English dual language learners. (For more details on the pretest, see Chapter III.)

In each pretest classroom, we collected information on the classroom and lead teachers. We identified someone in each setting (a setting point person, SPP) who answered questions about the children in the classroom (for example, age ranges, languages spoken in the children's homes). We also asked all lead teachers in the pretest classrooms to report information on their own background characteristics (such as education, experience, and demographic information) using a paper-and-pencil self-administered questionnaire (SAQ). (Appendix A contains a copy of this instrument.) Information from the caregiver SAQ provides some context for what we learned about the EDIT from the pretest.

### 1. Teachers

Of the nine teachers, eight worked full-time. All lead teachers were female, ranging in age from 30 to 48. Two teachers were Spanish-English bilingual. All the teachers had an associate's college degree or higher. Seven of the nine teachers had at least a bachelor's degree, and two of those teachers also had master's degrees. This was a highly educated group compared to, for example, data from the Family and Child Experiences Survey (FACES) 2009, which indicated that 35 percent of Head Start teachers had an associate's degree, 41 percent had a bachelor's degree, and 9 percent had graduate or professional degrees (Aikens et al. 2012). The teachers in the pretest had a range of early childhood teaching experience from 2 to 19 years, with a median of 10 years.

In addition to their advanced education, the teachers had professional development experiences that supported their ability to conduct and use assessments and plan lessons. All teachers reported professional development support specific to using assessments or evidence-based instruction. Their yearly training on these topics ranged from 5 to 40 hours a year. All

[23] Our final pretest visit involved testing revisions to the EDIT made after examination of the EDIT data across the first eight classrooms. The first eight visits occurred at the end of the program year, and the final visit took place in November, just a few months into the program year. We found that the EDIT could be used in the fall of the program year, as well as the spring, to identify areas for growth and capture strengths.

teachers also reported having mentors for these topics, and most teachers reported meeting with their supervisor about these topics a few times a year.

### 2. Communication with families

We asked teachers about the frequency of informal or formal communication with parents about how their children were doing. Five of the nine teachers reported speaking with parents at least every week, and some spoke with parents every few months or only at parent-teacher conferences.

### 3. Classroom assessment systems

Six of the nine classrooms used Teaching Strategies GOLD (TS GOLD; Teaching Strategies, Inc. 2011). Three classrooms used the Work Sampling System (WSS; Meisels et al. 2001).[24] Despite differences in these two systems (Table IV.1), we captured similar information on the EDIT across both systems. The most recent Program Information Report (PIR) data (2014) indicate that TS GOLD is the most frequently used preschool ongoing assessment system nationally. According to the 2013–2014 PIR data, more than half of all Head Start programs report using the preschool TS GOLD or its precursor, the Creative Curriculum Developmental Checklist. The WSS is used less frequently, with 4 percent of Head Start programs reporting using it.

Table IV.1. Two ongoing assessment systems

| | Ongoing assessment system | |
|---|---|---|
| **Characteristic** | **Teaching Strategies GOLD** | **Work Sampling System** |
| Areas measured by the assessment system | Social-Emotional<br>Language<br>Literacy<br>Cognitive<br>Math<br>Science and Technology<br>Social Studies<br>Arts<br>Physical<br>Each of the areas is comprised of objectives and dimensions. | Personal and social development<br>Language and literacy<br>Mathematical thinking<br>Scientific thinking<br>Social studies<br>Arts<br>Physical development<br>Each of the areas is comprised of objectives and dimensions. |
| Sources of assessment evidence | Anecdotes are based on observations and children's work collected throughout the day. | Anecdotes are based on observations and children's work collected throughout the day. Samples of children's work include "core work samples" collected at least once in each reporting period to examine progress across time.[25] |

[24] Several classrooms used assessments required by different entities (for example, district assessments three times a year), but the regular ongoing assessments in each classroom were either the TS GOLD or WSS.

[25] Our final pretest visit indicated that the most recent version of the WSS (fifth edition) no longer includes a separate portfolio.

| | Ongoing assessment system | |
|---|---|---|
| **Characteristic** | **Teaching Strategies GOLD** | **Work Sampling System** |
| Scoring | The teacher rates each child on each objective and dimension, which has a progression from simplest to most complex, based on expectations of children from birth to K.<br>Scores are levels from 1 to 9.<br>Each level indicates an age-level expectation. | The teacher rates each child on grade or age level objectives, comparing the child's performance to descriptive rubrics. Each learning objective is rated as "not yet," "in process," or "proficient."<br>Children can be scored using criteria for an earlier age range to meet developmental needs. |
| Age/grade level | Birth to 5 | Age 3 to grade 3 |
| Organization of data | Electronic database that organizes the data and produces multiple reports, graphs, and data displays at an individual child, domain, or classroom level, as well as by age or other subgroups within a classroom. | Electronic database that organizes the data and produces multiple reports, graphs, and data displays at an individual child, domain, or classroom level, as well as by age or other subgroups within a classroom. |
| Instructional activities | TS GOLD refers to the Creative Curriculum intentional teaching cards for individualization and differentiation of instruction. | This is intended to be used across curricula, so specific instructional activities for intentional teaching and individualization are not provided. |

## B. How well does the EDIT capture a range of quality in practices in pretest classrooms implementing ongoing assessment and individualizing?

This highly educated group of teachers and classrooms received above-average ratings on many of the items, allowing us to fine-tune the high end of many of the rubrics and also consider whether a checklist would suffice for some items.

The content of the items was revised throughout the pretest, and the scales were sometimes changed to evaluate whether we were capturing variance that was present in practice. When looking by stage of the conceptual model, most scores were on the high end of the scales throughout the pretest. Stages 1 and 2 of the conceptual model (planning and collecting the data) had overlapping items. For those two stages, 83 percent of the items show little variance: the items on a 7-point scale had means of 5 or greater. In stage 3 (interpreting data and formulating instructional decisions), 83 percent of the items on a 7-point scale had means of 5 or greater. Stage 4 (applying instructional decisions and individualizing) contained only one item, and the mean score of this item was above a 5. As we reflected on this, we considered that these centers were all recommended by researchers as quality implementers of assessment. As a group, their teachers were highly educated, and they all had some form of coaching or mentoring on assessment and instruction. Therefore, it was not surprising to find higher scores among this group.[26]

[26] On several items, all teachers received the maximum score. This may indicate an issue with items being too easy overall, rather than easy for this particular sample. We discuss this concern further in Chapter V in the section on items with limited variance.

### 1. Items with high mean scores

In this sample of teachers from recommended programs, we identified items which measure teacher practices that appear to be easy for teachers to do with quality. Our examination of assessment documents and observations indicate high scores on general assessment practices that are important for summative and performance-based assessment and that also can inform monitoring of progress. The most positive behaviors involved teachers' documentation and organization.

**Teachers' documentation.** Teachers collected most of the information during meaningful instructional activities (for example, writing how a child described a picture on the picture as the child dictated). Teachers' documentation of child performance or behavior was presented objectively, stating clearly what children did and/or said. The documentation often included information about the context. For example, an anecdote was:

> During small group, children were invited to play a sound sorting game. It was explained that there were several pictures we would sort. We went over all the pictures, then I explained that we would be matching those pictures that had the same beginning sound. CHILD was able to match spoon, spider, sun, snail, and star. (*Date was provided, and this was linked to particular objectives and scores on those objectives in the ongoing assessment system.*)

The teachers entered at least some documentation into the electronic ongoing assessment system regularly. This ranged from daily, to weekly, to whatever a teacher felt she needed to make a rating. Among those with more frequent data entered, the data entry times noted on the anecdotes suggested it was not unusual for these teachers to use their evenings and weekends to enter data. This timing of the data entry suggests that some teachers will be committed enough to the process to enter data regularly even in settings that do not offer time during the day to enter data. However, with a broad sample of teachers, we would expect that this is an item that will have strong variance in scores.

**Teachers' organization.** Both the TS GOLD and the WSS have an electronic database that organizes the information the teachers enter. Using this computer-based tool, teachers could access graphs or organized displays of data at the child, subgroup (most often by age), and/or classroom levels. In the TS GOLD, the displays of data included criteria for expected levels of performance at different ages. The WSS indicates the number of children who score at "not yet," "in progress," or "proficient" for each objective. Teachers reported using information from these data displays to inform their selection of goals and objectives for individual learning plans. We found that teachers usually selected areas of learning that were below age expectation or at the low end of performance bands for the child's current age during a previous summary assessment report (usually fall or winter). Several teachers used short-term individual plans and selected children who needed support in learning areas aligned with the lessons being taught that week. The observation of these patterns helped inform what questions to ask in the teacher interview when discussing the use of data.

**Teachers' instruction and individualization.** Teachers wrote individual plans and often drew on the use of small groups. Those using Creative Curriculum frequently used the activities

and teaching strategies described in the intentional teaching cards that accompany the curriculum (note: TS GOLD is aligned with Creative Curriculum).

Using information from all three data sources, we found that every teacher modified or adapted activities to support children's understanding and skill development. For example, in a video recorded observation, we saw that a teacher presented different versions of an activity based on the children's skills: some children traced their own name, others copied their name, and still others wrote their name from memory. However, except for one teacher, we did not hear evidence of intentionally selecting these strategies based on the information the teacher had about children's progress. We found that this difference in intentional use of data was not reflected well in the EDIT. Teachers received high scores on many of the items assessing the quality of data collection and organization of data, even when intentionality was not evident.

### 2. Evidence of a need for additions or changes to the measure

In reflecting on what we observed among teachers, we decided that we were not capturing some differences in practice, particularly related to how intentionally teachers gathered and used assessment for monitoring progress and selecting instructional strategies.

**Teachers' documentation.** Although the documentation that teachers provided included succinct anecdotal records that captured key information, several teachers described a process for obtaining those notes that was not efficient. It was not unusual for teachers to take copious running notes during free play time about different children. Teachers reported that not everything that they wrote down would be entered into the computer, because some of the notes were not relevant to what they needed to assess. When they went to enter these notes into the computer, they would look for a relevant area of learning for the observation. They sometimes would revisit their notes if they found that they did not have evidence of a child's skills or behavior in a particular area. We used this information to inform ratings about the efficiency of data collection, but thought that it also spoke to the intentionality in data collection that we were not capturing well in early drafts of the EDIT.

Before making our visits, we thought that perhaps a simple count of the number of anecdotal records and work samples would provide a good indicator of how often and how comprehensively teachers collect assessment information. However, we found that quantifying with a simple count would misrepresent the extent of the evidence collected. A single observation might be entered for multiple children even if a particular child's involvement did not demonstrate the skill, knowledge, or behavior being assessed. For example, a teacher collected an anecdotal record of a dramatic play interaction that recorded part of a conversation between two children. Rather than appearing just in the file of the two children conversing, the record appeared in the folder of all the children in that activity, even though it provided no information about the involvement of the other children in the activity beyond their presence. Occasionally, group activities were used as evidence of an individual child's proficiency. For example, in one case, a teacher took a picture of a graph that the class had created together, and that photo was included in every child's folder of documentation. That same teacher included the whole class counting to 20 as evidence of counting for individual children.

When we read through the available data, we found that evidence of a child's skill, knowledge, or behavior was limited in most areas to one or two observational records. When

questioned about what happened when they had no documented evidence, some teachers reported that they sometimes rated children based on their memory of what the child did. This provided additional support for the potential importance of assessing whether teachers had a schedule or plan for collecting assessment data across the domains of learning.

Most of the data that teachers collected were anecdotal records, with some photographs and work samples. Teachers collected most evidence about areas of learning that could be assessed using work samples or photographs. For example, teachers might collect multiple samples of children writing their names. Although checklists are less time-consuming to use than anecdotal records, teachers seldom used them to record their observations of children's skills, knowledge, or behavior. TS GOLD has a checklist for letter names and sounds that teachers typically used only at the reporting periods. The only other checklists we observed came from standard assessment tasks associated with state-funded preschool programs. These checklists were also completed only two or three times a year. We did not observe teacher-developed checklists. The limited use of checklists may indicate limited focus on progress across time. Checklists provide an easy way to examine progress. In addition, these behaviors may be another indicator of the level of intentionality in data collection. Use of a checklist requires some planning or preparation, but photos, work samples, and anecdotal records can be collected without any planning for data collection.

**Teachers' instruction and individualization.** When planning what skills and knowledge to address in individualizing, teachers focused on children's performance at a given time and how it compared to the criteria, to the performance of other children in the class, and sometimes to the performance of other children of the same age in the setting. None of the teachers articulated expectations for progress within a single reporting period, but rather looked at the most recent performance rating and where the child's performance fell in the age band designated for that child. For TS GOLD, the age band is a one-year period (3, 4, or 5 years old). Similarly, for the WSS, children are rated as "not yet," "in progress," or "proficient" in relation to indicators that have age-specific rubrics. When teachers reviewed the data for individual learning plans, they selected learning objectives on which children were performing at the low end of the age band or in the "not yet" or "in progress" categories rather than objectives on which children were performing in the high end of their age band or "proficient." When selecting an area for intentional teaching, some teachers also discussed considering the unit or theme that would be taught in the coming week, but others relied solely on the area of development where the child's performance was below the specified criteria and often below the performance of most of his or her peers. These observations of teacher practice led us to refine and add items to the EDIT to differentiate assessing progress versus performance. For example, we added this item: "The teacher monitors child progress in their area of individualization with at least 3 pieces of evidence for each child's area of individualization each quarter."

Teacher awareness of how to identify evidence-based or professionally recommended teaching strategies was limited. We probed this area to make sure that what we observed was capturing teacher practice. Teachers reported providing additional practice for children and/or selecting instructional strategies by choosing among two or three Creative Curriculum intentional teaching cards designated for a particular skill area or learning objective. Beyond that, teachers reported using colleagues, internet searches, and Pinterest as sources of ideas for activities and teaching strategies that addressed a particular skill or learning objective. A few

teachers mentioned drawing on strategies or curricular approaches from previous teaching positions. This suggests that the EDIT may provide important information about areas in which staff need more professional development or support.

After completing eight observations, reflections on teachers' use of assessment (Exhibit IV.1) in relation to the ratings on the EDIT suggested that our ratings were tapping some positive teacher assessment skills, but the items did not capture well the intentional use of assessment to inform instruction and individualization or to examine progress. For example, teachers could show evidence of providing prompts or environmental supports, but it was not clear that these supports were responsive to the data. We added indicators asking if the supports and prompts were documented and if they seemed responsive to the data. In our observations and study of children's assessment data, we did not find evidence of continued monitoring of individualized goals or of the effectiveness of any instructional strategy or intervention. Frequently, we would see an individualized plan for a child, but would not be able to locate any confirmation that the plan was actually implemented and could not find any documentation about how the child did on that learning objective after implementation of the plan. In relation to the conceptual model in Chapter I, the teachers enacted stage 4 (applying instructional decisions and individualizing) but did not continue the loop back to stage 1 (selecting the target and assessment method) by examining the effects of their attempt to individualize.

Exhibit IV.1. EDIT team reflections on nine teachers' use of assessment

| |
|---|
| The teachers' use of assessment was like following a recipe and including the specified ingredients for both assessment and intervention (identifying an assessment target, choosing a related instruction activity), but then putting it in to bake and not checking whether it turned out okay (whether the child's targeted skills, knowledge, or behavior increased). Teachers moved on to new recipes (sometimes new objectives, new interventions, or different children), rather than evaluating and refining the first recipe. They had no way of documenting whether they followed the recipe correctly (implemented the intervention well) and did not know if it was a healthy recipe (an evidence-based intervention). |

**Key findings**

Nine experienced classroom teachers were purposively selected to participate in the pretest.

- These teachers used the TS GOLD or the WSS ongoing assessment.
- The teachers' use of these systems was largely supported through training and mentors.

Overall, the teachers and classrooms received above-average ratings on the EDIT items. However, our reflections on what we saw and heard suggested the EDIT needed some revisions and additions to better capture intentionality in assessment and assessment of child progress in relation to instruction and individualization.

Items with high ratings (suggesting that these are easier practices for teachers to implement) included:

- Teachers regularly documented information about children objectively, and the documentation was collected during meaningful activities.
- The electronic ongoing assessment systems provided structured output that organized the data to facilitate interpretation.
- Most teachers planned individualized lessons, using materials that their curricula provided.

Evidence that led us to add items and revisions to better capture how intentionally teachers gathered and used assessment for monitoring progress and selecting instructional strategies included:

- Evidence of a child's skill, knowledge, or behavior was limited in most areas to one or two observational records, and the method for collecting these observational records often was inefficient and did not require planning.
- Teacher awareness of how to identify evidence-based or professionally recommended strategies was limited to use of curricular materials and colleagues.
- None of the teachers articulated expectations for child progress within a single reporting period.

Evidence of planning and intentionality in collecting and using data to inform instruction and individualization was lacking across classrooms.

In the next chapter, we describe how our pretest experiences shaped our procedures, as well as the EDIT instrument items and interview questions, as we developed the measure.

## V. LESSONS LEARNED FROM THE PRETEST

The goal of this work was to iteratively develop and test the EDIT measure. In this chapter, we describe what we learned about the process and the measure. We summarize themes that teachers shared with us during debriefs after our visit. We then describe challenges we experienced and changes made to our procedures and items during the pretest.

### A. Teacher debrief information: How did teachers feel about participation?

After our visits, we asked all teachers to debrief with a member of our research team by telephone. Seven teachers participated. The debrief discussions gave teachers an opportunity to comment on the research participation process, including the burden of collecting consents, gathering documents, and ease of video recording.

All who participated in a debrief call reported they would recommend participating in the study to colleagues. Teachers reported that the project looked overwhelming when they first received our materials, but all reported minimal burden after they got started.[27] Teachers found it easy to collect consent and typically spent one hour or less collecting documents for the document review.

The teachers' ability to video record themselves with the tablet we provided was a concern before we started the pretest. We sent written directions and offered each teacher technical assistance by telephone, but none of the teachers asked for assistance. Teachers reported that the tablets we provided were user-friendly. Most teachers compared the tablets to devices that they used at home. The resulting videos were of good audio and visual quality. Teachers provided us with four to eight videos ranging from 1 to 18 minutes, with an average total recorded length of 37 minutes per classroom. Some teachers practiced with the camera so, in total, it took about one hour to create the video recordings.

The on-site interviews with the lead teacher ranged from 45 minutes to one hour, with five of them taking a full hour. These interviews typically took place late in the school day or after school had ended. The interviews were most frequently in staff rooms outside of the classroom (for example, in a conference room). In total, over three weeks, teachers spent approximately three hours participating in activities related to the EDIT (Table V.1).

Table V.1. Estimated burden on teachers

| Activity | Time |
|---|---|
| Document review | 60 minutes |
| Recording videos | 60 minutes |
| Teacher interviews | 60 minutes |
| **Total time** | **3 hours** |

[27] We have since streamlined our materials to attempt to portray what is involved without overwhelming the teachers (Appendix B).

## B. What were challenges and changes to EDIT procedures?

During the pretest, we revised our data collection and administration procedures several times to more clearly communicate with teachers and obtain critical evidence from the different data sources to facilitate making EDIT ratings. Table V.2 provides examples of procedural changes we made.

Table V.2. Examples of changes made to procedures

| Procedure | Planned approach | Change made during pretest |
|---|---|---|
| Selection of focal children | Teachers complete checklist to help research team select two children. | Teachers self-select two children based on language and literacy performance but also continued to collect consent for all children in the classroom for video recording. |
| Describing and accessing documentation | Teachers are instructed to gather "assessment-related" documentation. | Teachers were instructed to gather "observation-" or "assessment-related" documentation and asked more about computer-based output. |
| Time period for documentation | Raters use worksheets to count data related to assessment targets in documentation gathered by teacher. | Emphasized a two-month time period for documentation and dropped the worksheet counts from EDIT procedures. |
| Focus of video recording | Video methods were planned to vary by teacher. | All teachers were instructed to use the same video method, with a telephone call offered to ensure the teacher understood the sequence. |

### 1. Selection of focal children

During the pretest period, we revised our procedure for selecting focal children. As noted in Chapter III, our initial plan was to use teacher ratings on the Oral Language & Comprehension measure (OL&C; Bradfield and McConnell 2013) to rank order consented children and then randomly select two focal children: one child with high ratings and one child with low ratings. However, two programs agreed to participate in the pretest shortly before their programs were ending in May. The time required to obtain consent from the families of all the children in each classroom, have the teacher complete the OL&C, and select the focal children had previously extended the period leading up to the site visits by three weeks, making it infeasible to complete before the end of the program year. Instead, we asked teachers to select one child with learning challenges in language and literacy and another child who was doing well in these areas. Teachers then only needed to obtain permission from the families of these two children to share data with us and to create the video recordings. We found that having the teachers nominate children with different levels of ability captured greater differences in child performance than our previous stratified random selection procedure and provided useful information about the teacher's interpretation of a child's skill, knowledge, and behavior.

This solution worked well for examining documentation and interpretation of data, but it was less successful for obtaining video recordings of typical teacher practice of individualization in instruction. Without consent from all the families, teachers were limited in their use of instructional grouping strategies in the video recordings. The video recordings from the teachers who selected the focal children themselves focused more on individual lessons with a child, sometimes in a separate room or during naptime, whereas the recordings from the teachers who collected consent on the entire class often included small group instruction.

For our final pretest visit, we requested that the teacher obtain consent for all children in the classroom to increase the chances of the teacher recording small group instruction, which was successful. We also requested that the teacher select the two focal children to capture greater differences in child performance.

### 2. Describing and accessing documentation

Initially, our materials requested that teachers gather "assessment-related" documentation, in addition to instructional documentation. We asked that teachers not change typical classroom practice, but rather share with us the documentation they were already collecting. We found that several teachers administered formal direct assessments that they normally would not have administered at that time of year. While these direct assessments had efficient methods for collecting and recording the data, they were only one source of information and did not affect the ratings. Based on how teachers were describing the ongoing assessments as "observations," we modified our language in materials to attempt to obtain only the data related to the ongoing assessment systems.

Another challenge to our work was accessing data stored in computer-based assessment systems. Many teachers printed anecdotal records for the focal children, but the set of records was not always complete. To protect the confidentiality of the other children in the classroom, at least one teacher excluded anecdotal records that included the names of children other than the focal child. To gain a full picture of how teachers use data to plan, we will need to obtain permission for all children in a classroom or provide a means for teachers to obscure the names of other children (thus increasing teacher burden).

It would also be beneficial to interview the teacher in her classroom, if possible. Interviews conducted in the classroom yielded richer information. In three of our visits, we were able interview the teacher in her classroom. Teachers were then able to jog their memory by providing examples in response to questions: they would look around the room at work samples on the walls or access notebooks or files kept in the classroom. In two of these classrooms, the teachers accessed the data entered into the system. We were able to observe information that was not available in the hard copy of the child's record. Because of varying levels of training, some teachers did not have experience with printing out the data and were not sure how to include information such as the objective or rating assigned to an anecdote or the different ways that they looked at the data when deciding how to individualize or what areas still needed data. For example, one center had just started with the new edition of the WSS, and the teacher was still struggling with how to access information in the system. In another example, one teacher demonstrated how she could look at the number of anecdotes that a child had in each domain and for each indicator. She used this information to decide what she would target in her observations. In at least one of the cases, the scores on a few items on the EDIT increased as a result of what the teacher was able to show us on the computer. This was true in a number of classrooms in relation to looking at data in different ways. Teachers did not consistently include the graphs and printouts of how they arranged the data when examining them on the computer. When we did receive the data displays, they sometimes were printed out by the administrator or education coordinator, or in one case, the teacher sought assistance in accessing data during our visit.

An additional consideration was access to the version of the classroom's ongoing assessment rubrics. Although TS GOLD is in its first edition, the WSS is now in its fifth edition,

and there were a number of changes between versions. Raters needed access to a copy of the assessment to examine how a teacher categorized a particular anecdote or other evidence in relation to the description and guidance provided by the assessment system. During the final pretest visit, the program used the most recent version of the WSS, and the assessment guidelines did not always align with the previous edition. This made it more difficult for us to evaluate whether the anecdotal records provided evidence aligned with the indicators. As development of the EDIT continues, we recommend asking programs to provide raters access to a copy of the program's assessment guidelines during the visit.

### 3. Time period for documentation

Before beginning pretesting, we developed detailed worksheets to collect information on the teachers' implementation of ongoing assessment, including counts of documentation records by child, domain, and assessment target. In most cases, we asked for all the data for the focal children for the two months (depending on the round of data collection) before the visit to see data across domains. Our intent was to look at patterns in data collection and how they might inform understanding of the teacher's use of data. For example:

- Did teachers prioritize some children or some domains more than others in collecting data?
- Did frequency of documentation differ across skills, knowledge, and behavior?
- Did teachers gather more information for children who were struggling and have very limited information about children who were doing well (particularly in the language and literacy and social-emotional domains) or vice versa?
- Did teachers collect information about some skills frequently across time and only have a single (or no) record about other skills or knowledge?
- Did teachers collect information about some skills in greater detail and use checklists or ratings for other skills?

Starting with our first visit, we realized how difficult it could be to collect some information, such as counts of anecdotal records for specific assessment targets that might help us identify patterns. This was particularly true in classrooms that had frequent child data collection. The first visits occurred in late spring, and the classrooms we visited had been recommended by early childhood researchers as moderate to high implementers of assessment. Therefore, we encountered classrooms that had a vast amount of documentation for each child, and even counting the number of anecdotal records within each developmental domain could consume much of our time in the setting.[28]

In addition, as mentioned in Chapter IV, we sometimes found that a single anecdotal record, photo, or checklist might be included in the documentation for many children. Furthermore, for a single child, the same anecdotal record might be repeated multiple times, providing evidence for multiple skills and domains. If we counted the number of records attributed to each domain and/or skill, we would overestimate the amount of evidence collected about a child. In

[28] Even when we asked teachers to limit the data they shared with us to a specified time period, they tended to give us all the data they had for the focal children. We will test our new procedures in the next phase of piloting.

subsequent visits, we sampled the previous two months of documentation and reviewed some data related to the child's entering skills and last summative report if available, as well as communications with families (usually only available for the triennial reporting periods). We found that the EDIT items were sufficient for helping raters identify pertinent information in the documentation.

### 4. Focus of video recording

During our iterative pretest process, we planned to test two approaches to the focus and frequency of the video recording (described in Chapter III as method 1, which had six video records across two children [three per child], and method 2, which had five video records for a single child). Although teachers were willing to video record and provided multiple recordings per child, we did not obtain videos that fully matched our requests. We frequently received video of small group or individual literacy instruction. Some teachers provided video of an end-of-year structured direct assessment. Other than the direct assessment documentation, we did not usually have evidence of teachers collecting data based on the video recorded lessons and interactions. Unfortunately, we were unable to either communicate clearly enough or motivate teachers to collect the information in sequence as requested. Because our verbal communications were filtered through the setting point person, we were not initially able to explain verbally to teachers about the sequence of activities we wanted video recorded. We phrased our written requests in different ways across classrooms to try to obtain the video recordings of interest, but we never received the exact sequence requested. Our revised written instructions did produce the exact number of videos requested; however, the content still differed slightly from what we had asked for. For the final pretest visit, we tried to speak directly with the teacher by telephone about our requested sequence; however, the teacher was not available for this call. A call with the teacher to walk through our video request would likely have produced a more positive outcome.

## C. What are the implications of the pretest for the feasibility of the measure if brought to scale and recommendations for future procedures?

In general, teachers reported that participating in the EDIT was not overly burdensome, taking teachers approximately three hours spread across several weeks to collect the necessary documentation, video record themselves, and participate in the teacher interview. Still, we are continuing to refine our communications with teachers to ensure that the materials needed are assembled without giving the impression that participation will be overwhelming. For example, a revised teacher request letter is in Appendix B. The original letter was long and detailed (three pages of information and a two-page tracking sheet). During the debrief conversations, teachers reported feeling overwhelmed after reading the original letter. Yet, teachers found participating in the EDIT process much easier than anticipated. We streamlined the text of the letter and used visual cues to communicate more succinctly. We also revised the language to avoid confusion about what we mean by "assessment." As noted earlier, some teachers interpreted "assessment" as referring only to direct assessments rather than to the ongoing data that they collected each day. Therefore, we referred to the data as "things that you collect and use to help decide how to support the learning of the 2 focal children."

Recommended procedures for future implementation of the EDIT include (1) gaining consent for all children while asking the teacher to select the two focal children; (2) refining

written materials to facilitate communication and to not overwhelm the teacher; (3) talking directly to the teacher at the outset whenever feasible to ensure understanding of the sequence of events that we would like video recorded and the data that we would like to review (providing time to troubleshoot how to best provide access to the information we need, including computer-based information); (4) requesting a copy of the program's edition of the ongoing assessment system; and (5) whenever possible, scheduling the teacher interview for a time when we can meet with the teacher in the classroom to provide the teacher easier access to a full range of materials.

### D. What were the contributions of different data sources?

In making ratings, we drew on different sources: (1) documentation, graphs, and data displays of child performance and progress, as well as instructional plans and individualization plans; (2) video recordings of assessments and instruction; and (3) teacher interview data. The data sources worked together to inform scores on a series of five holistic rubrics, two sets of analytic rubrics, two sets of ratings, and five checklists. As described in Chapter II, EDIT raters first reviewed documents, including child data, instructional and individualization plans, and graphs and data displays, and then assigned preliminary scores where possible. The raters noted missing information to look for in the observations and probes to include in the teacher interview. Raters then watched the video recorded observations and adjusted or assigned scores as needed. They again noted additional probes for the teacher interview. Finally, raters conducted the teacher interview and then finalized all scores.

Of the methods, the documentation and the teacher interview were the most critical for assigning ratings, but the video recording often clarified what was in the documentation or raised issues that were not evident in the written documentation. For example:

> A teacher wrote an individualized plan to use an alphabet book to help a child with letter recognition and video recorded two lessons with the alphabet book. We could see in the recording that the teacher was not keeping track of which letters the child named—even though this was a one-to-one interaction with the child. It became clear that the absence of documentation in the file was because it was not collected. In addition, the instruction we observed in those videos often focused on the meaning of the vocabulary word used to illustrate the sound of the letter of the alphabet, rather than on the letter itself. In watching the video, we noted that the child picked up some vocabulary words, but did not focus on the letters.

Across classrooms, we consistently found that the teacher interview was very helpful in understanding what we saw in the documentation and video observation. For example:

> One teacher collected work samples from the two focal children. The children had been asked to draw their favorite pet, as indicated by the words "Favorite Pet" written by the teacher on each paper. We could see obvious differences in visual motor and representation skills between the two children. The teacher had noted on the pictures (using quotation marks) how each child described his or her drawing. In one drawing, the child said, "Me and my dog walking to the park"; in the other drawing, the child just said, "Cat."

These work samples could provide information for the teacher about a child in the language, literacy, and fine motor domains. However, unless the teacher wrote a description or took a picture of this work sample and created an anecdotal record for it that she provided for us to review, we did not know how the teacher used this document until we asked her about it during the interview. Work samples are one of the many places where the interview allowed us to make a final rating that was more accurate than preliminary ratings based only on the documents or videos.

Overall, we recommend retaining all three data sources. This will allow us to understand each construct more accurately by approaching it from the different methodological perspectives that help capture both what teachers report and what they do (see Appendix D for additional data sources we considered).

## E. How did the EDIT items change during the pretest?

Two researchers from the EDIT development team rated each classroom. Researchers teamed together in different pairs across the observations. Each researcher rated independently, then the pair compared scores and discussed what was used as evidence for the rating, as well as any disagreements in ratings. When there were disagreements, they reached consensus. After the visit, the full EDIT development team met to discuss the disagreements and areas that were challenging to code.

Changes made to the EDIT instrument during the field period are described in this section and included:

- Refining the wording on the rubrics
- Refining the criteria within rubrics and ratings
- Changing some items to binary checklists (yes or no) or 4-point ratings rather than a 7-point rating
- Changing some ratings to the frequency of use, rather than tying them to the number of assessment targets
- Changing the flow of the questions in the teacher interview
- Adding questions and prompts to the teacher interview
- Collecting definitions and examples for key ideas in the items to share among the EDIT raters

### 1. Items with limited variance

EDIT items with limited or no variance included several items on which trained teachers scored highly when they knew how to implement and take advantage of the automated reports in the assessment system that the school used. Some of the affected items focused on criteria that were supported by features of the TS GOLD and WSS assessment systems (such as the organization of the data). Both systems can summarize and display information in a variety of ways, including at an individual child, domain, or classroom level, as well as by age or other subgroups within a classroom (Teaching Strategies GOLD Reports 2013). We revised item

response categories during the pretest to reflect the constrained variance on some items (Table V.3). For a number of these items, we changed the scale to a 1 to 4 scale (rather than 1 to 7) and asked raters to note whether the item was based on information gathered from all data sources.

During the final classroom visit of the pretest, two EDIT raters visited a program that had not provided training to the teacher on the WSS, the assessment system she was required to use in her classroom (although she did receive some mentoring on how to assess). We were able to observe more variance in the EDIT rating scales across items for this teacher. We were able to examine how some of the criteria on the lower end of the scales fit the classroom. Because we had limited variance in our earlier pretest visits, we tried both the 1 to 4 characteristic rating scale items and binary items when evaluating the documents, videos, and interview responses.

We found that items using a 1 to 4 rating scale better captured the variation in teachers' use of practices than yes or no ratings. Sometimes teachers would demonstrate a specified behavior for only one domain or only for children who were above average or only for children who were struggling. Sometimes they would demonstrate some of the described behavior, but not completely. For example, for the item "The teacher clearly communicates both strengths and challenges of the child with the family, providing evidence for each of these," we found that the teacher in the last visit sometimes communicated with families about children's strengths, but did not provide evidence based on ongoing assessments. Compared to the intent and scope of the item, the behavior described in the item was only "minimally characteristic" of the teacher. After the final visit, we changed some of the binary items to the 4-point rating scale. We recommend doing more testing with the rating supplement we tried during the final visit to decide if these items should be incorporated into the EDIT.

Table V.3. Changes for items with limited or no variance

| Item | Hypotheses for lack of variance | Action |
|---|---|---|
| 2d. Efficient method of assessment; does not take time away from instruction | Anecdotal records reflected naturally occurring events. Difficult to evaluate if teacher was missing instructional opportunities during documentation. | Examine in a more diverse set of classrooms; probe more about what the teacher is doing when anecdotal records and running records are being collected. |
| 2i. Assessments typically occur in a familiar context | Teachers implemented curriculum-based assessments as recommended. | Changed to a 1 to 4 rating, noting if the item is strongly characteristic. |
| 2g. Adequate attention to child during assessment | This is collected in the observation. Due to our method and the need for permissions, assessments were often 1:1 (during naptime or in another room) or pairs of children. | Maintained the item as a 7-point scale; however, will seek consent from all children in the classroom to facilitate video recording small groups. |
| 3f.Child strengths | Teachers tended to focus on writing down when a child was successful. | Changed to a 1 to 4 rating, noting if the item is strongly characteristic. |
| 3i. Objective documentation | Teacher education and professional development opportunity likely emphasized recommended general assessment practices. | Changed to a 1 to 4 rating, noting if the item is strongly characteristic. |
| 3j. Efficient documentation—no time lost to instruction | What is entered into online database is brief and to the point. | Changed to a 1 to 4 rating, noting if the item is strongly characteristic and reworded this item to emphasize intentional action. |
| 4p. Organization facilitates comparisons | TS GOLD and the WSS both offer a variety of types of graphs that teacher could use to examine group and individual differences (across domains and learning objectives). | Reworded this item to emphasize that the organization should facilitate understanding the current level of the child's skill compared to prior checkpoints/reporting periods. |
| 4g. Organization facilitates communication with families | TS GOLD and the WSS both offer a variety of types of graphs that teacher could use; programs also provide templates for reporting to families. | Changed to a 1 to 4 rating, noting if the item is strongly characteristic. |

## 2. Identified need to increase the EDIT's measurement of teachers' planning and evaluation of progress

Among the EDIT team members, as well as with the expert panel, we discussed the need for greater attention in the EDIT to how teachers plan and collect data that informs the effectiveness of instruction and individualization strategies. We wanted to capture the cyclical nature of the conceptual model, with each stage drawing on the previous stages and informing the next, repeating the cycle over and over. For example, at stage 1 (selecting the assessment target and method) a teacher may decide to assess whether a child can answer "How many are there?" for a group of 5 or fewer objects and plan to assess this at snack time using grapes. At stage 2 (implementing ongoing assessment), the teacher may put out a few grapes at a time and ask the child how many there are. Each time the child asks for more, the teacher puts out a different quantity and asks again. The teacher keeps track of how accurately the child tells how many and whether the child counted or just looked at the grapes and told how many. At stage 3 (interpreting data and formulating instructional decisions), the teacher looks at the data and notes that the child consistently named 1, 2, and 3 grapes right away, usually counted correctly when there were 4 grapes, and when there were only 5, counted the same grape twice, reporting 6

grapes. The teacher decided that she would group the objects in different ways when she put them in front of the child so that, for 5, she showed a group of 3 with a group of 2; or a group of 4 with a group of 1; or a line of 5. She planned to give this child 4 or 5 pretzel sticks at a time to give more practice on the numbers that were challenging. She created a checklist to keep track of when the child correctly counted the pretzels (returning to stage 1). She then implemented her instructional plan and documented what the child did on the checklist (stage 4 [applying instructional decisions and individualizing] and stage 2 [implementing ongoing assessment]). She then looked at how well the child correctly identified how many pretzel sticks there were and compared how she did compared to the day she counted grapes. If the child's was improving, she might continue using that strategy for the next week with different objects (such as counting bears, blocks, beads) and assess again looking for additional progress. If the child did not do well, she might select a different approach, such as showing the child how to move the object (for example, the grape) after it had been counted, in order to count each object only once.

As noted earlier, most assessment documentation we saw provided information on performance that was summarized at three times in the school year. Progress between reporting periods was not systematically examined, and progress/performance was examined in relation to the criteria that spanned the year. For many objectives, only one or two pieces of evidence (anecdotal notes, photos, work samples, or other documentation) were available, making it difficult for teachers to consider progress. In consultation with our experts, for our ninth pretest visit we included additional indicators in the EDIT rubrics to capture the intentionality with which teachers plan for and collect data to examine both child progress and the success of instructional strategies in supporting child progress (Table V.4). We tested both binary (N = 7) and rating scale items (N = 9), and we added to the criteria for 5 of the rubrics. We found that the rating scale items with a 4-point scale were the easiest to rate. The additional criteria made it more difficult to attain the highest rating on the revised rubrics. The two raters had exact agreement on the independent ratings of nearly all the new items.

Table V.4. Revisions made to the EDIT to increase measurement of teacher's intentionality

| Item | Example of revision |
|---|---|
| **Increased measurement of planning** | |
| 1 | This criterion was added to receive a score of 5: *At least one assessment target is represented in individualized plans for instruction.* |
| **Increased measurement of progress and the effects of individualization** | |
| 2j | A rating item was added that states: *Teacher monitors child's progress in area of individualization with at least 3 pieces of evidence (data points).* |
| 2k | A rating item was added that states: *When teachers individualize instruction, they collect information that allows them to see whether the child's current rate of progress is higher than his or her prior rate of progress.* |
| 2l | A rating item was added that states: *Teacher continues to periodically monitor child's progress in area of individualization even after first signs of improved progress.* |
| 3h | A rating item was added that states: *Documentation includes child's responses/performance during individual learning plan activities.* |

| Item | Example of revision |
|---|---|
| **Increased measurement of progress and the effects of individualization (cont'd.)** | |
| 4a | A binary item was added that states: *The teacher views/organizes the data to compare a child's performance to a developmental expectation or benchmark for growth.* |
| 4b | A binary item was added that states: *The teacher views/organizes the data to compare a child's progress to a developmental expectation or benchmark for growth.* |
| 4c | A binary item was added that states: *The teacher views/organizes the data to compare a child's performance to that of other children in the class.* |
| 4d | A binary item was added that states: *The teacher views/organizes the data to make it easy to understand the current level of the child's skill within reporting periods for individualization goals.* |
| 4e | A binary item was added that states: *The teacher organizes the data to look at a child's progress on individualized goals within a reporting period, looking at change based on at least 3 pieces of evidence (note: could be a comparison of 3 interim preliminary scores, work samples, or anecdotal records on a particular objective across time within a reporting period).* |
| 4h | A binary item was added that states: *The organization lets the teacher look at performance by class for one or more assessment targets at a single time point.* |
| 4i | A binary item was added that states: *The organization lets the teacher look at progress by class for one or more assessment targets across multiple time points.* |
| 4j | A rating item was added that states: *The organization lets the teacher look at performance by subgroup for one or more assessment targets at a single time point.* |
| 4k | A rating item was added that states: *The organization lets the teacher look at progress by subgroup for one or more assessment targets across multiple time points.* |
| 4l | A rating item was added that states: *The organization lets the teacher look at performance by domain for one or more assessment targets at a single time point.* |
| 4m | A rating item was added that states: *The organization lets the teacher look at progress by domain for one or more assessment targets across multiple time points.* |
| 4n | A rating item was added that states: *The organization lets the teacher look at performance by child for one or more assessment targets at a single time point.* |
| 4o | This criterion was added to receive a score of 7: *Teacher indicates when a new instructional strategy or individual learning plan is implemented.* |
| 5 | This criterion was added to receive a score of 5: *Teacher examines more than one data point in between checkpoints for at least one target per child (such as preliminary ratings).* |
| 5b | Additional guidance was provided. As an example, the criteria to receive a score of '7' specified: *Identifies when current rate of progress has accelerated beyond expectation even if current performance is still below age level. Identifies when current rate of progress has slowed or stopped even if current performance is above age level.* |
| 7 | Following criterion was added to receive a score of 7: *Organizes and reviews data (within reporting periods or checkpoints) to examine the effect of the individualization, and changes approach if the growth is not improving (that is, flat or negative).* |

For the final visit, we also added four items to the EDIT to address practices that we encountered or that we felt would be important (Table V.5). For example, we wanted to note whether teachers collected any documentation of child performance during the assessments they video recorded as part of participating in the EDIT. One item was specific to using paper data collection, so it was not applicable to our final visit.

Table V.5. Additional EDIT items

| Item | Example of revision |
|---|---|
| 2m | A rating item was added that states: Child's family or household members help collect assessment information. |
| 2n | A rating item was added that states: Teacher documents child behavior/performance or collects work sample (*based on video observations*). |
| 4r | A rating item was added that states: If using a teacher-developed system, the teacher files or enters data on a regular weekly basis. |
| 5c | A binary item was added that states: Teacher involves the family in interpreting and understanding the data. |

Based on a recommendation to obtain more information about the assessment context, we asked the teacher during the last visit for more information about program supports. Previously, we only captured these in the SAQ or heard about different supports during interviews, but we did not systematically probe this information. During the final visit, we asked the teacher about the amount of time available for assessing (including planning what information to collect and how, entering data, and reviewing data); time available and consultation available for planning instruction and individualizing; and availability of training in how to use the assessment. We noticed that a few responses during the interview did not reflect what the teacher recorded on the questionnaire. The teacher indicated more support for conducting and interpreting assessments than she reported in the interview. We may need to revise the wording in the SAQ to ensure that teachers clearly understand the questions they are asked about program support and/or probe further in the interview. This teacher was not available for the debriefing call, so we were unable to obtain additional information on the discrepancies we noted.

### 3. Use of "don't know" ratings

Nineteen items had at least one "don't know" by the EDIT raters. In some cases, the "don't know" was related to the timing of our data collection. In two classrooms, we were invited to come on the last day of the school year. The teachers had already sent home, securely filed, or destroyed some of the documentation that would have supported our ability to rate the teacher's practice. However, across the entire set of pretest classrooms, it was difficult to ensure that we had the documentation we needed to rate every item. We revised and refined how we worded our requests for documentation, but communicating clearly to the teacher what is needed without leading the teacher to create documentation that is atypical for that classroom remained a challenge. Different assessment systems and programs use different terms for the documentation we needed to rate the teacher's assessment practice. It was challenging to ensure that we had adequate and representative samples of teacher assessment behavior, instructional decision making, and individualization. The future users of the EDIT will need to make sure that the requests for information from teachers use terms that are familiar to teachers using that particular assessment system.

## 4. Summary

As the pretest progressed, we realized that the EDIT captured strengths in teachers' practices, but the high scores did not reveal some of the weaknesses we were seeing. We found that teachers' scores on the EDIT reflected general assessment practices rather than intentional progress monitoring to inform instruction and individualization. The original rubrics were weighted heavily toward valid and reliable documentation, organization of data, and some efforts to individualize instruction, but they did not capture the intentionality of assessment, or the use of data to inform decisions in the ongoing, cyclical manner represented in the EDIT conceptual model (Chapter I). As we made revisions, we thought carefully about what behaviors we would need to see as evidence that teachers used data to examine progress and to evaluate the success of instructional strategies and/or individualization plans, and we modified items accordingly.

**Key findings**

Seven teachers participated in debrief calls after the EDIT site visits.

- All reported they would recommend participating in the study to colleagues.
- Teachers reported that the project looked overwhelming when first receiving our materials, but all reported minimal burden after they got started.
- Teachers did not have difficulty operating the video equipment, and the resulting videos were of good audio and visual quality.

During the pretest, we revised our data collection and administration procedures, and we recommend the following practices:

- Gaining consent for all children in the class, while asking the teacher to select the two focal children.
- Refining written materials to facilitate communication and to not overwhelm the teacher.
- Talking directly to the teacher at the outset whenever feasible to ensure understanding of the sequence of events we would like video recorded and the data we would like to review.
- Requesting a copy of the program's edition of the ongoing assessment system for use during the visit.
- Whenever possible, scheduling the teacher interview for a time when we can meet with the teacher in the classroom to provide the teacher easier access to a full range of materials.

We recommend retaining all three data sources (document review, video recorded observations, and teacher interview) to understand each construct in the conceptual model and to be able to draw evidence from both what teachers say and what they do.

Key changes made to the EDIT instrument during the field period included refining the wording on the rubrics to clarify and specify concepts.

- Refining the criteria within rubrics and ratings, especially to capture measurement of teachers' planning and evaluation of progress.
- Changing some items to binary checklists (yes or no) or 4-point ratings rather than 7-point ratings in response to items with limited variance.

Key changes made to the EDIT teacher interview protocol during the field period included:

- Changing the flow of the questions in the teacher interview to facilitate conversation.
- Adding questions and prompts to the teacher interview, especially to capture measurement of teachers' planning and evaluation of progress.

The pretest was intentionally staggered in three waves to allow for a pause between visits to iteratively develop and refine the EDIT items and procedures. As expected, throughout the pretest, experiences in the field led us to refine our procedures and the EDIT instrument and interview to better capture the constructs identified in our conceptual model. A revised EDIT instrument and interview represent the culmination of what we have learned to date. In the next chapter, we propose further development of the EDIT.

## VI. FUTURE OF THE EXAMINING DATA INFORMING TEACHING MEASURE

In this chapter, we describe the proposal for refining the EDIT through continued pretesting. The chapter concludes with a discussion of the potential future use of the EDIT.

### A. Proposed plan for further pretesting

To further develop and refine the EDIT, we will visit additional classrooms with a wider range of assessment practice, including classrooms that are implementing both a standard tool (e.g., General Outcome Measures; GOMs) in addition to curriculum-embedded assessments. In these classrooms, we will specifically seek to refine our items related to:

- Teacher intentionality in planning and using assessment
- Interpreting child progress
- Use of information to inform next steps in both instruction and assessment

GOMs are standard tasks that are administered repeatedly to measure progress. They are based on indicators that have evidence of predicting important outcomes for children. For preschoolers, these include tasks like naming letters and numbers, looking at two quantities and telling which one is greater, and naming pictures (i.e., vocabulary growth). The most established GOMs in preschool are in language and literacy. In the last decade, more measures of preschool mathematics have been developed as well as a measure of social skills. The most commonly used preschool GOMs are the Individual Growth and Development Indicators (IGDIs; McConnell et al. 2002). Teachers can use curriculum-embedded assessments in combination with GOMs to measure development across all domains.

GOMs provide guidance about whether a child is making adequate progress or needs more help. Teachers who are trained to use GOMs make instructional changes and frequently monitor any children who are not meeting the adequate-progress criteria. They administer a GOM weekly or biweekly to see if the teaching strategy is helping the child to make greater progress. If progress is not occurring, they are trained to try a different evidence-based approach or more frequent or intensive instruction.

In our previous observations with the EDIT, we found that teachers were seldom intentional about their use of assessment to evaluate progress (as opposed to performance) or to evaluate the success of their instruction. Because teachers using GOMs are trained to use them to monitor progress, make instructional decisions, and evaluate the success of that instruction, we hypothesize that these teachers will be more intentional about their use of all assessments (including curriculum-embedded assessments) and would employ assessment data to inform and evaluate instructional decisions. These are areas of the EDIT that have new rubrics and ratings; therefore, visiting classrooms that make use of both GOMs and curriculum-embedded assessments may allow us to refine the new items.

## 1. Plans for continued pretesting

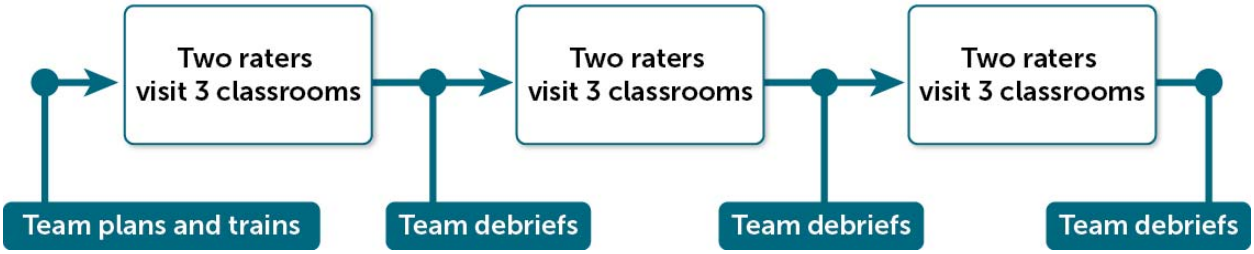
Applying the procedures described in Chapter II of this report, the CPM team plans to conduct three iterative rounds of data collection to further pretest the current EDIT instrument and teacher interview between October 2015 and March 2016. The primary goals of the continued pretest are (1) to implement and refine the newest EDIT rubrics and ratings, and (2) to continue to assess the overall feasibility of the EDIT's protocols, procedures, and materials.

Each round of data collection will involve one classroom in each of three different centers. Ultimately, the pretest will involve visiting nine centers and nine classrooms. Centers and teachers will receive a token of appreciation for their participation.

The centers will be chosen purposively based on (1) their willingness to participate, (2) their systematic use of a curriculum-embedded ongoing assessment system for individualization, and (3) their use of a standard tool (e.g., GOMs such as the IGDIs). We assume that centers using GOMs and curriculum-embedded assessments are likely to have teachers who implement intentional use of assessment. We would expect to find some variation in implementation based on our previous experiences in classrooms. A point of contact at each center will be asked to help us select a teacher in an English-speaking classroom within each center. To manage costs, we will attempt to recruit the classrooms in areas that are within driving distances of Mathematica's Washington, DC, and Princeton, New Jersey, offices when possible.

As in the earlier pretest, we will send two EDIT development team members to each classroom to score the EDIT, and the full team will debrief between each round of data collection (Figure VI.1). At the conclusion of the three rounds of data collection, the updated EDIT measure will be included in a final report. This report will also propose a plan for future development and testing of the EDIT.

Figure VI.1. Continued EDIT pretest activities




## 2. Benefits of continued pretesting

Continuing the EDIT pretest in an additional nine classrooms will allow us to refine the rubrics related to constructs that are likely critical to the overall process of using ongoing assessment to individualize instruction. We might find that increased emphasis in these areas also informs changes to other areas in the rubrics. For example, teachers who are intentional and targeted in the use of assessment to inform instruction and individualization might collect fewer overall pieces of assessment evidence or might implement other steps in the process in ways that differ from the practice that we observed in the pretest to date.

In addition, we could gather additional experiences and examples of assessment data to use in creating definitions and training materials for the EDIT. Clear definitions of terms and processes support the use of the assessment by a variety of assessors. Examples also help ensure that different raters interpret the rating rubrics in the same way. During the initial pretest, we began to collect definitions of key terms. For example, a familiar context was defined as "an activity setting that is familiar to the child." We also gathered examples of practices associated with criteria in the EDIT. For example, we added "Identifies rhymes in a finger-play or song" as an example of a target linked to a curriculum and meaningful outcome. We would continue to expand these materials with experiences in additional classrooms.

## B. Potential future uses

After more development, the EDIT could be used in the future for research and professional development. Research conference presentations of the literature review and measurement plan from the first phase of this project have already led to interest in the EDIT and requests to use it. It could inform the field in important ways.

### 1. Research

Research is the primary use for the EDIT at this time. We have a lot to learn about what teachers are doing to collect and use assessment data. Although the different indicators of quality included in the EDIT are professionally recommended, and some indicators have evidence of an association that supports the effectiveness of implementing that practice,[29] the research base on assessment practices is limited. Often, researchers examine assessment as a set of practices related to a specific assessment tool, and it is not possible to separate the practices from the assessment tool. More research is needed to determine which practices are most strongly associated with positive outcomes for children and to answer research questions such as:

- Can teachers implement the most supportive and effective assessment practices without implementing other assessment practices?
- Are the effective practices still meaningful and effective if done alone?
- What combinations of practices are needed for optimal benefits for children?
- If the most critical practices in the ongoing assessment cycle are in the interpretation, data-informed instructional planning, and execution of the instructional plan with examination of the success of that instruction, which of the indicators in the initial phases of the cycle (selecting a target and method, implementing assessment, organizing the data) need to be present for the teacher to successfully individualize for different children?

### 2. Professional development

The collection of evidence from the teachers for each of the EDIT rubrics and ratings provided a rich background for understanding the teacher's skills in collecting, organizing, interpreting, and using assessment information. With the support of the guidance provided in the rubrics, the strengths and weaknesses in the teachers' assessment practices were clear in the

[29] For example, technology supports that provide immediate recommendations to teachers have evidence of an association with child outcomes.

pretest. After the research has been conducted to more carefully examine the psychometric properties of the EDIT and the benefits of the practices involved, the EDIT could be used to evaluate teachers' use of assessment in programs. Researchers could use results from the EDIT to identify the most beneficial practices and develop professional development programs for teachers.

### 3. Evaluating and informing how programs support teachers in conducting and using assessment data

Any evaluative use of the EDIT would need to consider the context that supports successful implementation of ongoing assessment for informing instruction and individualization. The EDIT could be used to examine what program supports are needed for positive use of assessments to inform individualization. For example, one critical component of the use of assessment is the selection of an assessment that supports the teachers' understanding of the underlying constructs and how to help children make progress. Assessment systems offer different levels of support for the different steps in the process. Future research could provide information on features of assessments that are needed to support teachers in planning, collecting, and using assessment data. How much time do teachers need for professional development that focuses on implementation of the assessment system in use in their center? How much time do teachers need for planning, collecting, organizing, and reflecting on data to implement ongoing assessment for individualization? The EDIT could help in examining these and other questions related to the context for assessment.

## REFERENCES


Administration for Children and Families. "Head Start Performance Standards." 45 CFS Sec 1307.3(b)(2)(ii). *Federal Register,* November 9, 2011.

Aikens, N., E. Moiduddin, Y. Xue, L. Tarullo, and J. West. "Data Tables for Child Outcomes and Classroom Quality in FACES 2009 Report." OPRE report 201137b. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research & Evaluation, 2012. Retrieved from: http://www.acf.hhs.gov/sites/default/files/opre/data_tables_for_child_outcomes_and_classroom_quality_in_faces_2009.pdf.

Akers, L., P. Del Grosso, S. Atkins-Burnett, K. Boller, J. Carta, and B.A. Wasik. "Tailored Teaching: Teachers' Use of Ongoing Child Assessment to Individualize Instruction (Volume II)." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research & Evaluation, 2014.

Al Otaiba, Stephanie, Carol M. Connor, Jessica S. Folsom, Luana Greulich, Jane Meadows, and Zhi Li. "Assessment Data-Informed Guidance to Individualize Kindergarten Reading Instruction: Findings from a Cluster-Randomized Control Field Trial." *Elementary School Journal,* vol. 111, no. 4, 2011, pp. 535–560.

Atkins-Burnett, S., S. Monahan, L. Akers, J. Carta, B.A. Wasik, and K. Boller. "Tailored Teaching: Teachers' Use of Ongoing Child Assessment to Individualize Instruction (Volume I)." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research & Evaluation, 2014.

Atkins-Burnett, Sally, Emily Moiduddin, Louisa Tarullo, Margaret Caspe, Sarah Prenovitz, and Abigail Jewkes. "The Learning from Assessment Toolkit." Washington, DC: Mathematica Policy Research, March 2009.

Bagnato, Stephen J., Mary McLean, Marisa Macy, and John T. Neisworth. "Identifying Instructional Targets for Early Childhood Via Authentic Assessment: Alignment of Professional Standards and Practice-Based Evidence." *Journal of Early Intervention,* vol. 33, no. 4, 2011, pp. 243–253.

Bagnato, Stephen J., John T. Neisworth, and Kristie Pretti-Frontczak. *LINKing Authentic Assessment and Early Childhood Intervention: Best Measures for Best Practices.* Baltimore, MD: Paul H. Brookes Publishing Company, 2010.

Ball, C., and M. Gettinger. "Monitoring Children's Growth in Early Literacy Skills: Effects of Feedback on Performance and Classroom Environments. *Education and Treatment of Children,* vol. 32, 2009, pp. 189–212.

Bambrick-Santoyo, P. *Driven by Data: A Practical Guide to Improve Instruction*. San Francisco, CA: Wiley, 2010.

Black, P., C. Harrison., C. Lee, B. Marshall, and D. Wiliam. *Assessment for Learning: Putting It into Practice.* Buckingham, UK: Open University Press, 2003.

Black, P., C. Harrison, C. Lee, B. Marshall, and D. Wiliam. "Working Inside the Black Box: Assessment for Learning in the Classroom." *Phi Delta Kappan,* vol. 86, no. 10, 2004, pp. 8–21.

Bolt, Daniel M., Jim Ysseldyke, and Michael J. Patterson. "Students, Teachers, and Schools as Sources of Variability, Integrity, and Sustainability in Implementing Progress Monitoring." *School Psychology Review,* vol. 39, no. 4, 2010, pp. 612–630.

Bradfield, T.A., and S.R. McConnell. "Teacher Questionnaire of Language and Comprehension Skills Among Preschool Children." Minneapolis, MN: University of Minnesota, Center on Response to Intervention in Early Childhood, 2013.

Brewer, J., and A. Hunter. *Foundations of Multimethod Research: Synthesizing Styles*. Thousand Oaks, CA: Sage, 2006.

Buysse, V., E.S. Peisner-Feinberg, E. Soukakou, D.R. LaForett, A. Fettig, and J.M. Schaaf. "Recognition & Response." *Handbook of Response to Intervention in Early Childhood,* edited by V. Buysse and E. Peisner-Feinberg. Baltimore, MD: Paul H. Brookes Publishing, 2013, pp. 69–84.

Buzhardt, J., C. Greenwood, D. Walker, J. Carta, B. Terry, and M. Garrett. "A Web-Based Tool to Support Data-Based Early Intervention Decision-Making." *Topics in Early Childhood Special Education,* vol. 29, no. 4, 2010, pp. 201–213.

Buzhardt, J., C. Greenwood, D. Walker, R. Anderson, W. Howard, and J. Carta. "Effects of Web-Based Support on Early Head Start Home Visitors' Use of Evidence-Based Intervention Decision-Making and Growth in Children's Expressive Communication." *NHSA Dialog: A Research-to-Practice Journal for the Early Childhood Field,* vol. 14, no. 1, 2011, pp. 121–146.

Campbell, D.T., and D. Fiske. "Convergent and Discriminant Validation by the Multitrait Multimethod Matrix." *Psychological Bulletin,* vol. 56, 1959, pp. 81–105.

Carter, Deborah R., and Robert H. Horner. "Adding Functional Behavioral Assessment to First Step to Success: A Case Study." *Journal of Positive Behavior Interventions,* vol. 9, no. 4, 2007, pp. 229–238.

Chomat-Mooney, L.I., R.C. Pianta, B.K. Hamre, A.J. Mashburn, A.E. Luckner, K.J. Grimm, L. Wang, et al. *A Practical Guide for Conducting Classroom Observations: A Summary of Issues and Evidence for Researchers.* Report to the W. T. Grant Foundation. Charlottesville, VA: University of Virginia, 2008.

Connor, C., S. Piasta, B. Fishman, S. Glasney, C. Schatschneider, E. Crowe, and F. Morrison. "Individualizing Student Instruction Precisely: Effects of Child X Instruction Interactions on First Graders' Literacy Development." *Child Development,* vol. 80, no. 1, 2009, pp. 77–100.

DeBaryshe, Barbara D., Dana M. Gorecki, and Lori Mishima-Young. "Differentiated Instruction to Support High-Risk Preschool Learners." *NHSA Dialog,* vol. 12, no. 3, 2009, pp. 227–244.

Denzin, N.K., and Y.S. Lincoln. *The SAGE Handbook of Qualitative Research*. Thousand Oaks, CA: Sage, 2011.

Denzin, N. *Sociological Methods: A Sourcebook,* 2nd ed. New York: McGraw Hill, 1978.

Fuchs, L., S. Deno, and P. Mirkin. "The Effects of Frequent Curriculum-Based Measurement and Evaluation on Student Achievement, Pedagogy, and Student Awareness of Learning." *American Educational Research Journal,* vol. 21, 1984, pp. 449–460.

Fuchs, Lynn S., and Stanley L. Deno. "Paradigmatic distinctions between instructionally relevant measurement models." *Exceptional children,* vol. 57, no. 6, 1991, pp. 488–500.

Fuchs, L., and D. Fuchs. *What Is Science-Based Research on Progress Monitoring?* Washington, DC: National Center for Progress Monitoring, 2006.

Fuchs, Lynn S., Douglas Fuchs, Carol L. Hamlett, and Pamela M. Stecker. "Effects of Curriculum-Based Measurement and Consultation on Teacher Planning and Student Achievement in Mathematics Operations." *American Educational Research Journal,* vol. 28, no. 3, 1991, pp. 617–641.

Gallagher, L., B. Means, and C. Padilla. *Teachers' Use of Student Data Systems to Improve Instruction.* Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, 2007.

Goertz, M., L. Nabors Oláh, and M. Riggan. "From Testing to Teaching: The Use of Interim Assessments in Classroom Instruction." Philadelphia: Consortium for Policy Research in Education, 2009.

Good, R. H., J. Gruba, and R. A. Kaminski. "Best Practices in Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an Outcomes-Driven Model." In *Best Practices in School Psychology IV,* Volume 1, 4th ed., edited by A. Thomas and J. Grimes. Washington, DC: National Association of School Psychologists, 2001.

Good, Roland H., and Ruth A. Kaminski. "Assessment for Instructional Decisions: Toward a Proactive/Prevention Model of Decision-Making for Early Literacy Skills." *School Psychology Quarterly*, vol. 11, no. 4, 1996, pp. 326–336.

Greenwood, Charles R., Jay Buzhardt, Dale Walker, Waylon J. Howard, and Rawni Anderson. "Program-Level Influences on the Measurement of Early Communication for Infants and Toddlers in Early Head Start." *Journal of Early Intervention,* vol. 33, no. 2, 2011, pp. 110–134.

Grisham-Brown, Jennifer, Rena A. Hallam, and Kristie Pretti-Frontczak. "Preparing Head Start Personnel to Use a Curriculum-Based Assessment: An Innovative Practice in the 'Age of Accountability.'" *Journal of Early Intervention,* vol. 30, no. 4, 2008, pp. 271–281.

Hagans, Kristi S. "A Response-to-Intervention Approach to Decreasing Early Literacy Differences in First Graders from Different Socioeconomic Backgrounds: Evidence for the Intervention Validity of the DIBELS." *Assessment for Effective Intervention,* vol. 34, no. 1, 2008, pp. 35–42.

Hamilton, L., R. Halverson, S. Jackson, E. Mandinach, J. Supovitz, J. Wayman, C. Pickens, E. Sama Martin, and J. Steele. "Using Student Achievement Data to Support Instructional Decision-Making." NCEE 2009–4067. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, September 2009.

Heckman, J., and L. Raut. "Intergenerational Long-Term Effects of Preschool—Structural Estimates from a Discrete Dynamic Programming Model." National Bureau of Economic Research working paper 19077. Cambridge, MA: NBER, May 2013.

Hojnoski, Robin L., and Kristen N. Missall. "Monitoring Preschoolers' Language and Early Literacy Growth and Development." *Young Exceptional Children,* vol. 10, no. 3, 2007, pp. 17–27.

Hosp, John L., and Scott P. Ardoin. "Assessment for Instructional Planning." *Assessment for Effective Intervention,* vol. 33, no. 2, 2008, pp. 69–77.

Kim, Do-Hong, and JaneDiane Smith. "Evaluation of Two Observational Assessment Systems for Children's Development and Learning." *NHSA Dialog,* vol. 13, no. 4, 2010, pp. 253–267.

Kim, Do-Hong, Richard G. Lambert, and Diane C. Burts. "Evidence of the Validity of Teaching Strategies GOLD® Assessment Tool for English Language Learners and Children with Disabilities." *Early Education & Development,* vol. 24, no. 4, 2013, pp. 574–595.

Lambert, Richard G., Do-Hong Kim, and Diane C. Burts. "Using Teacher Ratings to Track the Growth and Development of Young Children Using the Teaching Strategies GOLD® Assessment System." *Journal of Psychoeducational Assessment*, vol. 32, no. 1, 2014, pp. 27–39.

Landry, S., J. Anthony, P. Swank, and P. Monseque-Bailey. "Effectiveness of Comprehensive Professional Development for Teachers of At-Risk Preschoolers." *Journal of Educational Psychology,* vol. 101, no. 2, 2009, pp. 448–465.

Landry, Susan H., Paul R. Swank, Jason L. Anthony, and Michael A. Assel. "An Experimental Study Evaluating Professional Development Activities Within a State-Funded Pre-Kindergarten Program." *Reading and Writing: An Interdisciplinary Journal,* vol. 24, no. 8, 2011, pp. 971–1010.

Lee, J., W. Grigg, and P. Donahue. "The Nation's Report Card: Reading 2007." NCES 2007–496. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, September 2007. Available at http://nces.ed.gov/nationsreportcard/pdf/main2007/2007496.pdf. Accessed July 15, 2009.

Lo, Ya-yu, Chuang Wang, and Sherry Haskell. "Examining the Impacts of Early Reading Intervention on the Growth Rates in Basic Literacy Skills of At-Risk Urban Kindergarteners." *Journal of Special Education,* vol. 43, no. 1, 2009, pp. 12–28.

Luze, Gayle J., and Kere Hughes. "Using Individual Growth and Development Indicators to Assess Child and Program Outcomes." *Young Exceptional Children,* vol. 12, no. 1, 2008, pp. 31–41.

Marsh, J., J. Pane, and L. Hamilton. *Making Sense of Data-Driven Decision-Making in Education: Evidence from Recent RAND Research.* (OP-170). Santa Monica, CA: RAND Corporation, 2006.

Martin, T. "Teachers' Beliefs About Classroom Assessment and Their Selection of Classroom Assessment Strategies." *Journal of Research and Reflections in Education*, vol. 6, no. 2, December 2012, pp. 104–115.

McConnell, Scott R., Mary A. McEvoy, and Jeffrey S. Priest. "'Growing Measures for Monitoring Progress in Early Childhood Education: A Research and Development Process for Individual Growth and Development Indicators." *Assessment for Effective Intervention,* vol. 27, no. 4, 2002, pp. 3–14.

Meisels, S., D. Bickel, J. Nicholson, Y. Xue, and S. Atkins-Burnett. "Trusting Teachers' Judgments: A Validity Study of a Curriculum-Embedded Performance Assessment in Kindergarten–Grade 3." *American Educational Research Journal,* vol. 38, no. 1, 2001, pp. 73–95.

National Association for the Education of Young Children, Division for Early Childhood of the Council for Exceptional Children, and National Head Start Association. "Frameworks for Response to Intervention in Early Childhood Education: Description and Implications." Review draft. Available at http://www.naeyc.org/files/naeyc/RtI_in_ECE_Frameworks_DRAFT_FOR_REVIEW_6-27-12.pdf. Accessed December 11, 2014.

National Center for Student Progress Monitoring. "Common Questions for Progress Monitoring." 2012. Available at http://studentprogress.org/progresmon.asp#2. Accessed December 11, 2014.

National Research Council of the National Academies. *Mathematics Learning in Early Childhood: Paths Toward Excellence and Equity,* edited by Christopher Cross, Taniesha Woods, and Heidi Schweingruber. Washington, DC: National Academies Press, 2009.

Patton, Michael Quinn. *Qualitative Research and Evaluation Methods,* 3rd ed. Thousand Oaks, CA: Sage, 2002.

Peisner-Feinberg, Ellen, and Virginia Buysse. "Early Education." In *Handbook of Response to Intervention in Early Childhood,* edited by V. Buysse and E. Peisner-Feinberg. Baltimore, MD: Brookes Publishing Company, 2013, pp. 121–142.

Peisner-Feinberg, Ellen, and Virginia Buysse. "The Role of Assessment Within Response to Intervention in Recognition & Response: A Model of Response to Intervention to Promote Academic Learning in Early Education." In *Handbook of Response to Intervention in Early Childhood,* edited by V. Buysse and E. Peisner-Feinberg. Baltimore, MD: Brookes Publishing Company, 2013, pp. 69–84.

Pretti-Frontczak, Kristie, Stephen Bagnato, Marisa Macy, and Dawn Burger Sexton. "Data-Driven Decision Making to Plan Programs and Promote Performance: Applying Best Professional Practice Standards." In *Childhood Intervention: Programs and Policies for Special Needs Children,* three volumes. New York: Praeger, 2011.

Ritchie, Jane, and Liz Spencer. "Qualitative Data Analysis for Applied Policy Research." In *The Qualitative Researcher's Companion,* edited by A. Michael Huberman and Matthew B. Miles. Thousand Oaks, CA: Sage, 2002.

Roehrig, Alysia D., Staci W. Duggar, Louisa Moats, Marsha Glover, and Brian Mincey. "When Teachers Work to Use Progress Monitoring Data to Inform Literacy Instruction: Identifying Potential Supports and Challenges." *Remedial and Special Education,* vol. 29, no. 6, 2008, pp. 364–382.

Rowe, M., S. Raudenbush, and S. Goldin-Meadow. "The Pace of Vocabulary Growth Helps Predict Later Vocabulary Skill." *Child Development,* vol. 83, no. 2, 2012, pp. 508–525.

Sikka, A., J.L. Nath, and M.D. Cohen. "Practicing Teachers Beliefs and Uses of Assessment." *International Journal of Case Method Research & Application,* vol. 19, no. 3, 2007, pp. 239–253.

Teaching Strategies, Inc. "Teaching Strategies: GOLD®." Washington, DC: Teaching Strategies, Inc., 2011.

U.S. Department of Health and Human Services. "Improving School Readiness and Promoting Long-Term Success: The Head Start Roadmap to Excellence." Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Head Start, January 2010. Available at http://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/class/Head_Start_Roadmap_to_Excellence.pdf. Accessed December 12, 2014.

VanDerHeyden, Amanda M., Carmen Broussard, Melanie Fabre, Jessica Stanley, Jaclyn Legendre, and Raegan Creppell. "Development and Validation of Curriculum-Based Measures of Math Performance for Preschool Children." *Journal of Early Intervention,* vol. 27, no. 1, 2004, pp. 27–41.

VanDerHeyden, Amanda M., Patricia A. Snyder, Carmen Broussard, and Kerrie Ramsdell. "Measuring Response to Early Literacy Intervention with Preschoolers at Risk." *Topics in Early Childhood Special Education,* vol. 27, no. 4, 2008, pp. 232–249.

Wasik, B., A. Hindman, and A. Jusczyk. "Using Curriculum-Specific Progress Monitoring to Determine Head Start Children's Vocabulary Development." *NHSA Dialog: A Research-to-Practice Journal for the Early Intervention Field,* vol. 12, no. 32, 2009, pp. 257–275.

Webb, E., D. Campbell, R. Schwartz, and L. Sechrest. *Unobtrusive Measures: Nonreactive Measures in the Social Sciences.* Chicago: Rand McNally, 1966.

Ysseldyke, Jim, and Daniel M. Bolt. "Effect of Technology-Enhanced Continuous Progress Monitoring on Math Achievement." *School Psychology Review,* vol. 36, no. 3, 2007, pp. 453–467.

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

# APPENDIX A

# TEACHER SELF-ADMINISTERED QUESTIONNAIRE (SAQ)

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

Mathematica Ref. No. 40158.B33

MATHEMATICA
Policy Research

# Examining Data Informing Teaching (EDIT)

## Caregiver Questionnaire

*March 13, 2014*

## ABOUT THIS QUESTIONNAIRE

**This questionnaire is an important part of a larger study supported under a contract from the U.S. Department of Health and Human Services, Administration for Children and Families. The overall purpose of the Examining Data Informing Teaching (EDIT) project is to understand the ways teachers use assessments to individualize instruction for preschool children. Participation in this project is voluntary.**

**This form requests information about your child-care setting and your background and experience. The information will be used for research purposes only and will be kept confidential to the extent allowed by law. Your answers to these questions will not be shared with your employer. Your name will not be attached to any information you give us. Please note that pages are double-sided, and the questionnaire is 3 pages. You may skip any question you do not wish to answer.**

**Most of the questions can be answered by marking an "X" in the box. For a few questions you may be asked to write in a response.**

1 ☐ 2 ☐ X 3 ☐

**Thank you very much for your help.**

## A. STAFF DEVELOPMENT

**A1. Please record today's date:**

|__|__| / |__|__| / |2|0|__|__|

**A2. How many hours a year do you attend staff trainings about assessment or evidence-based instructional practices?**

|__|__|__| HOURS

**A3. How often do you have one-on-one supervision meetings or group supervision meetings about assessment or evidence-based instructional practices?**

**MARK ONE ONLY**

0 ☐ Never

1 ☐ Once a year

2 ☐ A few times a year

3 ☐ Every 2 months

4 ☐ Once a month

5 ☐ Twice a month

6 ☐ Once a week

7 ☐ More than once a week

n/a ☐ Not applicable

**A4. Is there someone who mentors you in your classroom, that is, someone who observes your teaching on a regular basis and provides feedback, guidance, and training about assessment or evidence-based instructional practices?**

1 ☐ Yes

0 ☐ No

## B. COMMUNICATION WITH PARENTS

**B1. How often do you talk to parents about how their children are doing on a formal or informal basis?**

**MARK ONE ONLY**

0 ☐ Never

1 ☐ Only at parent-teacher conferences

2 ☐ Every 2 or 3 months

3 ☐ Once or twice a month

4 ☐ Once or twice a week

5 ☐ Daily

**B2. How often do you hold formal parent-teacher conferences with parents about individual children?**

**MARK ONE ONLY**

0 ☐ Never

1 ☐ Once a year

2 ☐ Twice a year

3 ☐ 3 times a year

4 ☐ 4 or more time a year

## C. EMPLOYMENT

**C1. Are you currently working in this early childhood setting full or part-time?**

**MARK ONE ONLY**

1 ☐ Full time

0 ☐ Part time

**C2. Counting this school year, how long have you worked in your current early childhood setting?**

|__|__| YEARS |__|__| MONTHS

**C3. Counting this school year, how long have you worked in your current classroom?**

|__|__| YEARS |__|__| MONTHS

**C4. Please indicate your role(s) at this early childhood setting.**

**MARK ALL THAT APPLY**

1 ☐ Owner

2 ☐ Director

3 ☐ Lead Teacher

4 ☐ Assistant Teacher

5 ☐ Teacher

6 ☐ Administrative Assistant

7 ☐ Other role (*please specify*)

_____

## D. EDUCATION AND EXPERIENCE

**D1. Do you currently hold a Child Development Associate (CDA) credential?**

1 ☐ Yes

0 ☐ No

**D2. What is the highest level of education you have completed?**

**MARK ONE ONLY**

0 ☐ High school diploma or GED

1 ☐ College course(s) without a degree

2 ☐ Associate's degree

3 ☐ Bachelor's degree

4 ☐ Master's degree

5 ☐ Education specialist or professional diploma based on at least one year of course work past a Master's degree level

6 ☐ Doctorate

7 ☐ Other (*please specify)*

_____

**D3. In what field did you obtain your highest degree?**

**MARK ALL THAT APPLY**

1 ☐ Child development or developmental psychology

2 ☐ Early childhood education

3 ☐ Elementary education

4 ☐ Special education

5 ☐ Other (*please specify)*

_____

**D4. How many college courses have you completed related to child development and/or assessment?**

|__|__| CHILD DEVELOPMENT COURSES

|__|__| ASSESSMENT COURSES

**D5. Including this year, how many years have you worked with preschool aged children?**

|__|__| YEARS

## E. DEMOGRAPHIC INFORMATION

**E1. Are you…**

1 ☐ Female

2 ☐ Male

**E2. In what year were you born?**

|__|__|__|__| YEAR

**E3. Please indicate languages you speak fluently.**

**SELECT ONE OR MORE**

1 ☐ English

2 ☐ Spanish

3 ☐ Other (*please specify*) _____

**E4. Are you of Spanish, Hispanic or Latino origin?**

1 ☐ Yes

0 ☐ No

**E5. What is your race?**

**SELECT ONE OR MORE**

1 ☐ White

2 ☐ Black or African-American

3 ☐ Asian

4 ☐ American Indian or Alaskan Native

5 ☐ Native Hawaiian or other Pacific Islander

**Thank you for your participation. If you have any questions about this questionnaire or the EDIT project, please call Felicia Hurwitz, at (609) 945-3379.**

**Please return this questionnaire in the envelope provided. If you no longer have the envelope, please mail this questionnaire to:**

**Mathematica Policy Research**

**Attn: Receipt Control – Project 40158**

**P.O. Box 2393**

**Princeton, NJ 08543-2393**

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

# APPENDIX B

# LETTER TO TEACHERS

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

MATHEMATICA
Policy Research

**Shannon Monahan**
*EDIT Project Director*

P.O. Box 2393
Princeton, NJ 08543-2393
Telephone (609) 275-2207
Fax (609) 799-0005
www.mathematica-mpr.com

DATE

Dear [TEACHER NAME]:

Your classroom has been selected to be part of an important project called the Examining Data Informing Teaching (EDIT) project. Your participation will help the U.S. Department of Health and Human Services, Administration for Children and Families better understand ways in which teachers collect information about the children in their classrooms and how teachers use that information to help children learn.

We look forward to visiting your center and meeting with you on DATE. To help you get your materials collected and organized for our visit, we ask that you prepare as follows:

| | |
|---|---|
| First<br>**Contact Felicia Hurwitz via email** (fhurwitz@mathematica-mpr.com) **or by phone (609-945-3379) to schedule a time to discuss upcoming activities** | **On this brief call, we will:**<br>Review procedures to help you prepare for our visit<br>Provide technical assistance<br>Answer any questions you have about the project |
| Next<br>***Select 2 Focal Children*** | Select 1 child performing well and 1 child who has some challenges in language or literacy. Please use the following list to select from among children who have permission to participate.<br>**DO CONSIDER SELECTING:** [INSERT INITIALS OF CONSENTED CHILDREN]<br>**DO NOT CONSIDER:** [INSERT INITIALS OF NON-CONSENTED CHILDREN] |
| Week of DATE<br>***Collect Round 1 Videos*** | Four separate videos<br>Child doing well (in a small group or individually):<br>1. Working on language or literacy activity/lesson<br>2. Activity where you are collecting information about a child's knowledge or skills in language and literacy<br>Child with challenges (in a small group or individually):<br>3. Working on language or literacy activity/lesson<br>4. Activity where you are collecting information about a child's knowledge or skills in language and literacy |
| Week of DATE<br>***Collect Round 2 Videos*** | Two separate videos:<br>Child doing well:<br>1. Working on a language or literacy activity/lesson<br>Child w/challenges:<br>2. Working on a language or literacy activity/lesson |



An Affirmative Action/Equal Opportunity Employer

MEMO TO:
FROM:
DATE:
PAGE:

| | |
|---|---|
| Between Today and DATE<br>Gather Documents | From the previous two months, all the things that you collect and use to help decide how to support the learning of the 2 focal children including any observations, assessments, and plans for instruction.<br>Examples of what some teachers use:<br>Anecdotal records, photos, checklists, charts, graphs<br>Portfolios of children's work<br>Family reports, class reports<br>Lesson plans ; individualized learning plans<br>Instructional sequence, unit plans, yearly themes or projects, schedules for observing or collecting information about children |
| DATE<br>EDIT Team Visit | On DATE, 2 researchers will visit your center to:<br>Review documents you gathered<br>Review the videos you recorded<br>Speak with you for one hour to learn more about how you use assessment data to plan strategies for children in your classroom<br>Ask you to complete a brief paper questionnaire |

Enclosed in this package, you will find an iPad mini and a tripod with instructions for recording videos with the focal children. The researchers will collect the iPad mini from you when they arrive for the one day site visit.

Please note that if your program allows, we will provide you with a $75 gift card and your child care program with a $50 gift card in appreciation for your participation in this important study.

In addition, these tips and notes may be helpful as you prepare for our visit:

| Preparation | Tip | Note |
|---|---|---|
| Select 2 Focal Children | Think about the children in your class and select one child performing well in language and literacy and another child who is struggling. | Please do not select children who do not have permission to participate.<br>**DO CONSIDER SELECTING:** [INSERT INITIALS OF CONSENTED CHILDREN]<br>**DO NOT CONSIDER:** [INSERT INITIALS OF NON-CONSENTED CHILDREN] |
| Videos with Focal Children | Create a designated recording space with limited background noise, and position the iPad so that ***only*** you and the consented child(ren) are visible in the video.<br><br>For each of the videos, please focus on language or literacy learning, but otherwise follow your typical classroom practice.<br>Please only collect information in the ways that you typically do in your classroom.<br>Confirm that your videos saved to the iPad after each recording session. | Do not include pictures of other adults or non-consented children – even in the background – in your videos. |

MEMO TO:
FROM:
DATE:
PAGE:

| Preparation | Tip | Note |
|---|---|---|
| Gather Documents | Gather the information that you collected about the 2 focal children from the 2 months prior to our visit and also any plans for instruction from those 2 months. | We understand that paperwork varies by classroom, and we don't want to add to your paperwork, so please **only share documentation that you already have on-hand.** We will return all documentation at the end of our visit; names will be hidden in any copies/photographs made of the documentation. |

If you have any questions, please contact me by email at SMonahan@mathematica-mpr.com or by phone at 609-275-2207. Thank you in advance for your time and consideration.

Sincerely,

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

# APPENDIX C

# TRAINING PLAN

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

To extend development of the EDIT beyond the pretest stages, a training session would be necessary to proceed with more extensive data collection. Trainees should have knowledge of evidence-based early childhood instructional practices in the language and literacy and social-emotional domains and experience conducting observations in early childhood classrooms. Trainers would include members of the EDIT development team.

Before the training, the EDIT team will submit the training agenda, a list of training materials, and a plan for evaluating performance via a certification process to ACF. After approval of these documents, we will submit draft training materials, including manuals and a description of training exercises, to ACF for documentation before the start of training.

The EDIT development team will produce a data collection manual to serve as a reference manual for EDIT raters. A critical portion of this manual will be an item-by-item description of the instrument with examples and definitions of terms developed based on our pretest experiences, along with scoring procedures. The trainers will review contents of the manual with trainees during training. Using data collected in the pretest supplemented by materials prepared by the EDIT development team, the observation training will involve presentations of high- and low-quality practices related to implementation of ongoing assessment and individualization of instruction to help observers recognize quality practices. It also will include multiple opportunities to practice with the measure by rating documents, watching and rating videos, and listening to recordings of teacher interview, followed by group discussion.

Goals for training are to ensure that trainees obtain the following skills:

- Fluency in the goals of the project, and rationale for the selected sample so that trainees can respond to questions from center staff and parents as needed
- The ability to reliably complete the EDIT measure, including an understanding of expectations, availability of materials, and the terminology used by these different ongoing assessment systems
- A thorough understanding of the measure development process, so that raters can participate in frequent debriefs and comfortably adjust to changes in the EDIT or its procedures in the middle of data collection
- Proficiency in administration of the semi-structured teacher interview and the introduction to, and distribution of, teacher SAQs
- An understanding of the necessity of, and procedures for, maintaining strict confidentiality and data security, including protecting paper copies of documents with personally identifiable information (for example, securing paper copies in folders so these are not visible and keeping these folders with them in the field [not left in cars or hotel rooms] and then shipping these back to Mathematica's office using a national shipping service and tracking numbers as soon as the field period in that location is complete)

- The ability to complete documentation accurately
- The ability to work efficiently and well with team members, the Setting Point Person (SPP), and teachers at each center

Members of the EDIT measure development team will lead in-person training sessions. Our plan for an interactive and scaffolded training will feature a variety of formats to maximize trainees' active engagement and learning. The training format will include the following:

- **Formal presentations to convey information, accompanied by handouts for future reference** (for example, PowerPoint presentations describing the project goals).
- **Discussions to help trainees process information and to provide an informal opportunity for trainers to clarify information and correct misunderstandings.** (For example, when trainees justify their scoring of a particular item on the EDIT it can highlight trainees' thought processes about the item and can provide an opportunity to reinforce or correct an approach.)
- **Review sample documents, video recordings, and recorded teacher interviews from the pretest to learn how to complete the EDIT.** Discussions between trainers and trainees to justify master scores to particular items will follow. We will provide corrective feedback and explanation to trainees as needed.
- **Interactive exercises to maintain engagement and enable trainees to enact portions of their role in a supportive learning environment** (for example, role playing mock teacher interviews, or answering frequently asked questions).
- **Field observations to practice in realistic conditions.** After trainees have practiced and discussed completing the EDIT with videos, interviews, and mock documents, they will apply the rubrics to live observations in local classrooms. A member of the EDIT development team will accompany trainees to each classroom during the field practice. Following the practice observation, the trainees will have an opportunity to discuss their scores with a trainer in small groups and understand the differences between their scores and those of the trainer.

Training will culminate in a certification process designed to reflect field procedures for the site visit and EDIT rating processes.[30] Trainees will be required to rate using the EDIT rubrics and a standard set of materials designed to reflect EDIT procedures. The standard set of materials will include mock documents, as well as videos and recordings of interviews from the pretest. After certification, a trainee will fall into one of three categories:

1. **Fully certified.** Trainees who pass certification will be permitted to conduct EDIT site visits.

[30] Reliability thresholds on the EDIT would be determined in consultation with ACF.

2. **Provisionally certified.** Trainees who are close to certification but have not yet met the level required to conduct the measure reliably may receive provisional certification status at the discretion of the EDIT training team. Provisionally certified trainees will be required to practice and conduct a second certification exam.

3. **Not certified.** Trainees who do not clearly demonstrate required skills (for example, exhibit low reliability) and proficiencies by the end of the training will not remain on the project.

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

# APPENDIX D

# ADDITIONAL METHODS CONSIDERED

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

## APPENDIX D. ADDITIONAL METHODS CONSIDERED

In addition to the methods that we implemented in the EDIT, we also considered two other data collection methodologies: (1) testing pedagogical content knowledge (PCK), and (2) using a standard pedagogical task. Here, we describe the advantages and disadvantages of incorporating these approaches into a measure of teacher assessment and individualization practices. The EDIT development team evaluated the strengths and weakness of each of these methods and ultimately decided not to include either of these methodologies in the EDIT. For more details on the EDIT measurement plan, please see Atkins-Burnett et al. 2014.

### PCK with scenario probes

PCK questions capture the intersection of a teacher's knowledge of child development, assessment, and instruction. Questions and probes can be designed to isolate what teachers know from the contextual circumstances that can affect what they do. PCK questions could incorporate scenarios with probes about decision-making processes that would capture what teachers know about ongoing assessment and how to interpret and use information gained from assessment. The scenario questions can be multiple-choice or open-ended (such as asking teachers to write short paragraphs reacting to a scenario and presenting differentiated instruction suggestions).

PCK questions could potentially be developed to examine constructs from each of the four stages in the conceptual model. Teachers' knowledge of child development, pedagogy, and assessment contributes in different ways to each stage in the model. For example, teachers could be asked to identify a task that would provide evidence of the fine motor development of a three-year-old. In another example, teachers could be shown a sample of a child's dictation about a picture and the child's age and asked if this sample of the child's language is evidence that the child is performing below age level, at age level, or beyond typical development for age.

### Standard pedagogical task

Standard pedagogical tasks can be used to gauge teachers' ability to use ongoing assessment data. For example, a standard task could ask a teacher to examine another teacher's documentation of a child and identify what areas of learning are assessed with that learning sample, discuss hypotheses and interpretations about the child's development, and describe ways to tailor instruction for that child. Standard tasks could also involve asking a teacher to explain how she would group children for differentiated instruction.

### Advantages of PCK questions and standard pedagogical tasks

Both PCK questions and standard tasks have some advantages. They both tap into what teachers know and think about using ongoing assessment for individualization. If a teacher is not implementing ongoing assessment (as indicated by his or her scores on the EDIT rubrics), then PCKs and standard tasks could help clarify whether it is a lack of knowledge of what to do or if the issue is more likely related to other factors, such as demands on the teacher's time or the teacher's beliefs about the importance of using data to inform instruction.

Because a given teacher's classroom might not include children at varying levels of performance, document reviews and observations might not enable us to compare a teacher's ability to work with children who have diverse backgrounds and abilities. PCK questions (with scenario probes) and standard tasks are not limited to the children in a teacher's classroom. PCK questions and standard tasks can provide a point of comparison across teachers by examining what each teacher knows about working with children at varying levels of performance, although these options cannot gauge whether and how a teacher implements that knowledge.

### Challenges with incorporating PCK questions and a standard pedagogical task

In consultation with ACF and the expert panel, the EDIT development team identified three major drawbacks that prevented us from incorporating PCK questions and a standard task into the EDIT to date. First, teachers interpret PCK scenario probes based on hypothetical, rather than actual, assessment results. This prevents them from drawing on the contextual knowledge they would otherwise have if they were looking at data from their own classrooms. Second, creating the items and field-testing their reliability and validity can be costly and poses challenges. In addition to creating multiple items for each dimension, a PCK test or standard task could require creating multiple equivalent forms so that teachers cannot share answers and are not given the same question or task repeatedly, which would bias results. Furthermore, the EDIT is currently being developed for use with various assessment systems. Teachers are trained to look at data in different ways, depending on the assessment tool they are using in their classrooms. Even when teaching universal tasks (for example, teaching children how to write letters), teachers are applying a set of rules that are particular to the assessment system they are using. Teachers might know how to implement the assessment system that they are using, but might not be able to apply guidelines or rubrics from new systems. Although these methods could be a useful complement to the EDIT, they are not sufficient to understand what teachers actually do. Therefore, they are not included in the EDIT.