



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



OPRE Report #2016-103
Updated November 2016



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 #2016-103

Updated November 2016

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

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

<http://www.acf.hhs.gov/opre/research/project/early-childhood-teachers-use-of-progress-monitoring-to-individualize>

Contract Number: HHSP23320095642WC/HHSP23337041T

Mathematica Reference Number: 40158

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, 2016.

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.



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, Sangeeta Parikshak 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. Other staff members from the Office of Head Start 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 (see page v), 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; Grace Lim and Kara Conroy who attended site visits. Patricia Ciaccio and Donovan Griffin provided editorial support. Laura Sarnoski and Brigitte Tran designed the figures in the report. Jill Miller and Dorothy Bellow 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 (ACF) 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

Leslie Nabors Oláh
Educational Testing Service

Linda Broyles
Southeast Kansas Community Action
Program

Sheila Smith
Columbia University

Virginia Buysse
American Institutes for Research

Patricia Snyder
University of Florida

Lynn Fuchs
Vanderbilt University

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

CONTENTS

ACKNOWLEDGMENTS.....	v
EXPERT PANEL MEMBERS.....	vii
OVERVIEW.....	xiii
EXECUTIVE SUMMARY.....	xiv
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 multimethod measure focusing on two domains.....	11
1. The need for a multimethod 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.....	17
3. Teacher interview.....	17
C. Summary of the measure.....	18
1. Number and types of items.....	18
D. Administration procedures.....	37
1. Step 1: Selecting focal children.....	37
2. Step 2: Teacher activities before the raters' visit.....	37
3. Step 3: EDIT rater(s)' visit.....	38

III	THE MEASUREMENT DEVELOPMENT PROCESS.....	41
	A. Recruitment	41
	B. Pretest timeline and data collection procedures.....	42
	1. Recruitment and consent	42
	2. Selecting focal children	44
	3. Teacher activities before the raters' visit	44
	4. Raters' visit.....	47
	5. Self-administered questionnaire	47
	6. Teacher debrief	47
	C. Refining the instrument and its procedures.....	47
IV	PRETEST.....	49
	A. Sample: Who participated in the pretest?	49
	1. Teachers	49
	2. Communication with families	50
	3. Classroom assessment systems	50
	B. How well does the EDIT capture a range of quality in practices in pretest classrooms implementing ongoing assessment and individualizing?	51
	1. Items with high mean scores	52
	2. Evidence of a need for additions or changes to the measure	53
V	LESSONS LEARNED FROM THE PRETEST	58
	A. Teacher debrief information: How did teachers feel about participation?	59
	B. What were challenges and changes to EDIT procedures?.....	60
	1. Selection of focal children	60
	2. Describing and accessing documentation	61
	3. Time period for documentation	62
	4. Focus of video recording.....	63
	C. What are the implications of the pretest for the recommendations of future procedures?.....	63
	D. What were the contributions of different data sources?	64
	E. How did the EDIT items and interview protocol change during the pretest?	65
	1. Items with limited variation.....	66
	2. Identified need to increase the EDIT's measurement of teachers' planning and evaluation of progress.....	68
	3. Use of "don't know" ratings	72
	4. EDIT instrument conventions.....	72

5. Interview protocol	72
6. Training materials.....	73
7. Summary.....	73
VI NEXT STEPS AND THE FUTURE OF THE EDIT.....	75
A. Recommended next steps for further work	75
1. Challenges to future work	77
2. Sample size.....	77
3. Recommended procedures for pilot test.....	81
4. Validity evidence	82
B. Potential future uses.....	84
1. Research.....	85
2. Professional development.....	85
3. Examining and informing how programs support teachers in conducting and using assessment data	85
REFERENCES.....	87

APPENDICES

APPENDIX A: EDIT INTERVIEW PROTOCOL

APPENDIX B: TEACHER SELF-ADMINISTERED QUESTIONNAIRE (SAQ)

APPENDIX C: LETTER TO TEACHERS

APPENDIX D: TRAINING PLAN

APPENDIX E: TRAINING DEFINITIONS AND EXAMPLES

APPENDIX F: SAMPLE TRAINING AGENDA

APPENDIX G: 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	20
II.2.	Proposed schedule for site visit	39
III.1.	ounds of data collection for the pretest	41
III.2.	Depiction of differences in pretest data collection, by round	43
III.3.	Consent collection during pretest.....	44
IV.1.	Two ongoing assessment systems.....	51
V.1.	Estimated burden on teachers	59
V.2.	Examples of changes made to procedures.....	60
V.3.	Changes for items with limited or no variation	68
V.4.	Revisions made to the EDIT to increase measurement of teacher's intentionality.....	70
V.5.	Additional EDIT items.....	71
VI.1.	Sampling selection criteria options ^a	79
VI.2.	Commonly used ongoing assessments as reported in the 2014–2015 Program Information Report	81
VI.3.	Advantages and disadvantages of constructs considered for validating the EDIT	83

EXHIBITS

II.1.	Examples of documentation.....	16
II.2.	Definitions of EDIT item types.....	19
IV.1.	Study team reflections on some pretest teachers' use of assessment.....	56

FIGURES

I.1.	Conceptual model for curriculum-embedded approaches	6
II.1.	Multimethod measure model.....	16
II.2.	EDIT procedures	37
II.3.	Method for the focus and frequency of video recordings.....	38
II.4.	Use of data sources to assign scores	40
III.1.	Alternative methods for the focus and frequency of video recordings.....	46

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 multimethod approach in gathering evidence with checklists, ratings, and rubrics. 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 18 classrooms were selected to participate in the pretest. Supervisors or research partners nominated teachers who collected and used ongoing assessment information regularly. During the first nine visits, there was limited variability in EDIT ratings with teachers and classrooms receiving high ratings on many of the initial EDIT items. After revising some EDIT items, we found greater diversity in the teacher ratings during our next nine visits. We found evidence of high scores on some general assessment practices. For example, teachers regularly documented information about children objectively and collected the documentation during meaningful and authentic classroom activities. Because only 18 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 for teachers to receive a high rating on the EDIT.

We also identified EDIT items for which teachers generally received lower ratings. Those items allowed us to examine variability in 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 in the early versions of the EDIT. This was particularly true for constructs related to how intentionally teachers gathered and used assessment for monitoring progress and selecting instructional strategies. We revised the EDIT to improve our coverage of these constructs by adding additional criteria to our items after our first nine visits. With the more nuanced criteria, there was more variability with some teachers exhibiting the behaviors at the high end of the rubrics in our next nine visits.

Throughout the pretest, we revised our data collection 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, added indicators to better capture levels of teachers’ intentional planning and evaluation of progress, and added conventions to the instrument (such as “Not Applicable” responses). To test the psychometric properties of the revised version, we recommend visiting additional classrooms with a wider range of assessment practices, including greater diversity in teacher skill and in the type of assessment system used. After additional testing and adaptation, the EDIT measure could be used to identify beneficial practices for collecting and using ongoing assessment data to individualize instruction. In addition to its use as a research tool, with additional testing and adaptation, the EDIT shows promise as a management tool, supporting the use of ongoing assessment for individualization.

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.¹ This report includes the results of a pretest study in 18 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 multimethod approach where the rater gathers 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 six rounds of iterative pretesting for the EDIT measure. After each round, in consultation with ACF, the EDIT study team met and discussed challenges that arose and proposed changes to improve the procedures and EDIT items to obtain more precise 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. Several times during the pretest, we shared the EDIT with the project’s expert panel, and incorporated their feedback.

Teachers in 18 classrooms were purposively selected to participate in the pretest. Supervisors or research partners nominated teachers who collected and used ongoing assessment information regularly. Of the 18 teachers, 14 had at least a bachelor’s degree, and 8 of those teachers also had master’s degrees. The teachers in our pretest had early childhood teaching experience that ranged from 2 to 32 years. Ten of the 18 classrooms used Teaching Strategies GOLD (TS GOLD; Teaching Strategies, Inc. 2011). Six classrooms used the Work Sampling System (WSS; Meisels et al. 2001). One classroom used Tools of the Mind (Bodrova and Leong 2007). One classroom used a school-developed system for ongoing assessment. The teachers had at least some training, and most had coaches or mentors to support the use of these assessment systems.

¹ The EDIT was formerly called the Tool for Tailored Teaching (T3).

Teachers in the pretest sample consistently received high scores on several EDIT items. We found evidence of some high quality general assessment practices that suggests it is relatively 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.² For some teachers, the electronic ongoing assessment systems (for example, TS GOLD) provided structured output that organized the data to make interpretation easier. Most teachers planned individualized lessons, using curricular materials aligned with the objectives that they assessed. During the first half of the pretest, 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 revised EDIT items for the second half of the pretest by adding measurement of intentionality and increased the criteria at the higher ends of the rating scales. These revised items resulted in greater variation in scores.

During the first half of the pretest, we identified some items that allowed us to examine the lower end and the middle of the rubrics, but did not allow us to examine the high end of the rubric (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 true 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, initially, 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 teachers during the first half of the pretest did not 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.

² Meaningful activities support learning in the goals and objectives being assessed.

- **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. Only one teacher described using assessment data to determine whether or not an instructional approach was helping.

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, added indicators to better capture measurement of teachers' intentional planning and evaluation of progress, and added conventions to the instrument (such as "Not Applicable" rating options). 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. We expanded our measurement of planning, intentionality, progress, awareness of evidence based instructional strategies, and continued monitoring. With these modifications, we found greater variation in teacher's EDIT scores during the second half of the pretest (including evidence of intentionality, planning, and attention to progress within reporting periods).

Only 18 classrooms were assessed with the EDIT during the development phase. More testing is needed. To further examine the psychometric properties of the EDIT, the study team recommends 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. We refer to this expanded testing effort as a pilot test. The four primary goals of an expanded pilot test would include (1) testing the EDIT with a broader set of classrooms, looking for sensitivity to differences in practices and teacher backgrounds; (2) evaluating the reliability (including both internal consistency and inter-rater reliability) and validity of the measure; (3) evaluating different approaches to scoring the EDIT, and (4) expanding operational definitions. In addition, we recommend gathering samples of assessment data to use in creating training materials for broader use of the EDIT. A future pilot test should also collect more information to examine how characteristics of the classroom, program, or context (for example, class size, classroom composition) are related to differences in how teachers implement assessment. A pilot test should also examine how teachers' assessment and individualization practices vary by characteristics of assessment systems.

To examine convergent validity, it would be ideal to include an additional measure in the pilot test. However, given the absence of any other measure of ongoing assessment practices, it will be challenging to identify other measures to use to examine the validity of the EDIT. Although some other factors (such as observed instructional quality) are likely to be related to assessment practices, we would expect only a weak to moderate correlation with those measures. This suggests the need for careful selection of potential related constructs and a large sample size to ensure adequate power to detect relationships, particularly if there is interest in examining subgroup differences (such as differences between assessment systems).

The EDIT is specifically designed to help researchers understand how teachers use ongoing assessment to inform and modify instruction when necessary. Use of evidence-based practices

when using ongoing assessment data to individualize instruction may help teachers more effectively meet the needs of children. In addition to its use as a research tool, with additional testing and adaptation, the EDIT shows promise as a management tool supporting the use of ongoing assessment for individualization.

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 rates 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 includes four sections: (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 the way 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 (ACF 2016).

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 (Peisner-Feinberg and Buysse 2013; Buysse et al. 2016; Carta et al. 2015; Greenwood et al. 2013, 2015, 2016; Hamilton et al. 2009; McConnell et al. 2015; 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; Piasta 2014). 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 (Buysse et al. 2016; 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 six years, the Office of Head Start 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 summarize individual child information at least three times a year (Administration for Children and Families 2016).³

Despite the importance of using ongoing assessment data to guide instruction—and the Head Start program requirements to do so—information is sparse on the way 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; Classen and Cheatham 2015; Fuchs and Fuchs 2006; Hamilton et al. 2009; Marsh et al. 2006; Zweig et al. 2015). 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 we⁴ 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 group of experts who were consulting on this project to recommend seminal work before 2002 and articles 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

³ Head Start regulations require a child’s first developmental screener be conducted within 45 days of enrollment.

⁴ Throughout this report, “we” refers to the study team.

the study 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 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)	n.a.	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.

RCT = randomized controlled trial; QED = quasi-experimental design; SCD = single-case design.

n.a. = not applicable.

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 was available about the areas critical for the successful implementation of ongoing child assessment to individualize instruction. However, the existing literature provided some recommendations for how ongoing assessment should be used for individualization and also described the range of activities we were 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; Macy and Bagnato 2013; Pretti-Frontczak et al. 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 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 those data 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 three things: (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, recent research on ongoing assessment to individualize instruction was limited. However, it provided important insights that helped inform a conceptual model that focused on understanding the use of ongoing assessment for individualized instruction and potential approaches to measuring those constructs.

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.

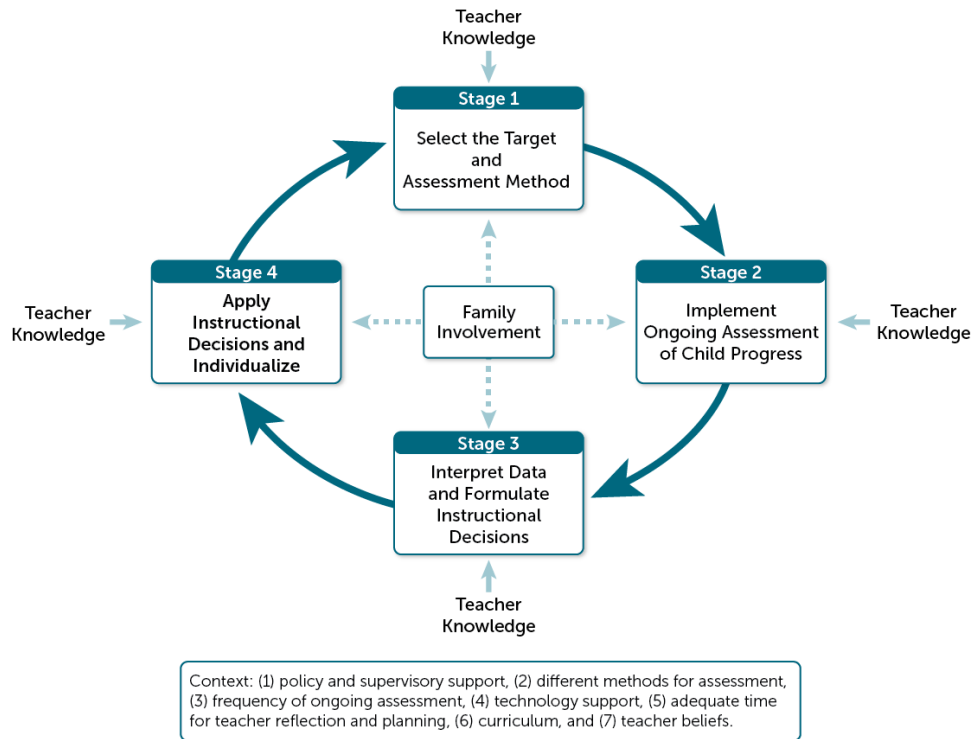
Box 1. Approaches to ongoing assessment in preschool

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 above expected level of performance for age.

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 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. The teacher can 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 need to ensure that they implement the instructional strategies with fidelity, that the content is correct, and that 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 three 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. 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 multimethod 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 six 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 multimethod 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, reviews of manuals for curriculum-embedded assessments, and experiences during iterative pretesting. Together, this information helped the study 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, including logistics of implementation and cost of development and ongoing use. Overall, the project's goal was to develop the EDIT so that researchers could easily use it in diverse settings to yield high quality data.

This chapter first describes the rationale behind the EDIT's multimethod 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 multimethod measure focusing on two domains

In this section, we first discuss the need for a multimethod approach. We then provide an overview of the EDIT's three data sources.⁶

1. The need for a multimethod approach

Experts in research design and methodology strongly recommend using a multimethod 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 semistructured 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 might organize data for interpretation (although the teacher would need to use that feature to receive credit on the EDIT).

Given the complexity of teachers' use of ongoing assessment for individualization, a multimethod 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 the way 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 those data to individualize instruction.

Several factors can affect the processes involved in teachers' use of ongoing assessment data. It is important to recognize the factors that may affect the use of assessment data, and the data sources in the EDIT that inform understanding of these factors:

- **Teacher knowledge (interview and, to a lesser extent, the document review, and observation).**⁷ Understanding what teachers think as they complete assessments and individualize instruction sheds light on the way 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 because of 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 (Appendix A). During the pretest, however, the team relied primarily on a brief 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 B). 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 included four requirements: (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 multimethod approach and the amount of time it might take to train raters to learn and implement the different components.

Burden. The study team used a multimethod 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 and 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 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 semistructured 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 study 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 multimethod 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. 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. We asked each teacher to use his or her judgment in selecting two children: one child who is doing well in language and literacy and one child who has difficulty or struggles in one of these areas. Overall, teachers collected the most evidence about language and literacy so we hypothesized that it would be easiest to select children based on those domains.

3. Initially targeting two domains

All the domains in the Head Start Early Learning Outcomes Framework (Administration for Children and Families 2015) are important for understanding child progress and, therefore, are relevant to this study. However, several constraints prohibited exploring all domains in the initial

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.

phase of measurement development. Perhaps most prominent was the burden on teachers, classrooms, and raters when data are collected in many domains.

Instead of imposing this burden, the expert panel suggested pretesting only two domains: (1) language and literacy, and (2) social and 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, 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 of the observer training¹⁰ and the video recording 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 reports employing 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 or focus on another domain—for example, one that is the focus of a professional development effort. 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.

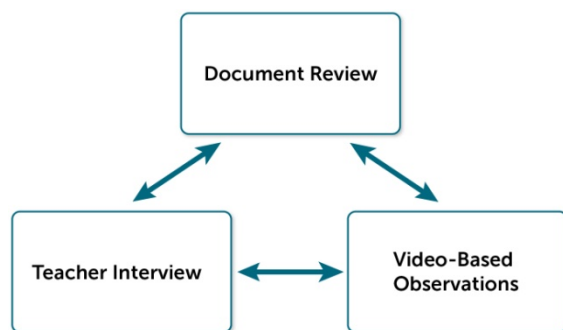
B. Overview of the three methods

As Figure II.1 shows, the multimethod 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 visit¹¹ to review the documents, watch the video, and interview the teacher. Next, we discuss each of the EDIT methods in detail.

¹⁰ Raters needed to be trained to recognize the presence of evidence based or professionally recommended instructional practices targeted to a child’s individual goals. To do this for all domains would have been overwhelming for raters.

¹¹ Depending on the amount of data provided by the teachers in the documents and videos, the site visit during the development phase required four to six hours in a single day.

Figure II.1. Multimethod measure model



1. Document review

Teachers provide documentation for the two focal children: one performing well in language and literacy and another facing challenges.¹² 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

<p>Assessment documents</p> <ul style="list-style-type: none"> • 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 <p>Instructional documents</p> <ul style="list-style-type: none"> • 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. The

¹² The team experimented with focusing on ratings of children's social and emotional development or language and literacy in selecting the focal children. Because teachers tended to have the most data points for language and literacy, we transitioned to asking teachers to select focal children on the basis of language and literacy. During the document review, video-based observations, and teacher interview, however, the raters were attentive to children's development in all domains, including the social and emotional domains.

rater makes preliminary EDIT ratings based on the document review. The document review can provide information for constructs in all four stages of the conceptual model, and it is particularly useful for stage 1 (selecting the assessment target and method) and stage 2 (implementing ongoing assessment).

2. Video-based classroom observations

During the two weeks before the raters' visit, the lead teacher in each classroom uses a tablet to video record approximately six activities: two instructional language, literacy, or social skills activities/lessons per focal child (working with each focal child either individually or in a small group) and at least one assessment activity per focal child (in which the teacher collects information about the child's knowledge or social skills).

Video-recorded observations provide critical information about how teachers conduct assessments and use the data to inform their instructional practice. This provides further evidence for what is examined in the document review. 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.¹³

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 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. A cover sheet on the front of the interview protocol highlights key topics the interviewer should cover (Appendix A).

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 about the documents and video-recorded observations (for example, "Using [CHILD] as an example, please walk me through your process for collecting information, reviewing the information, and then using the information to inform instruction"). Before the interview starts, the rater organizes these questions to align with stages in the conceptual framework. These questions are nonstandard probes that vary based on the documents and video contents. For example, reflective

¹³ 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.

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 tasks recorded on the video. The rater may use additional probes, but needs to take care to avoid leading teachers and skewing the results. 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.
- 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.
- 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 can 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. We then present the items' relation to the conceptual model, and present the EDIT measure.

1. Number and types of items

The EDIT instrument consists of four holistic rubrics, three sets of analytic rubrics, five sets of ratings, and two checklists, for a total of 71 items (Exhibit II.2 and Table II.1).

Exhibit II.2. Definitions of EDIT item types

Definitions of EDIT item types	
Rubrics	<p><i>Rubrics</i> 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 7) but are more elaborate than ratings, often including multiple criteria that must be met before an item can earn points on the rating scale.</p> <p><i>Analytic rubrics</i> 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).</p> <p><i>Holistic rubrics</i> 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.</p> <p>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).</p>
Ratings	<p><i>Ratings</i> are scales that take measurements along a continuum (for example, 1 to 4 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.</p>
Checklists	<p>Raters use <i>checklists</i> 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.</p>

Note: For additional information on measurement options selected for the EDIT, see Atkins-Burnett et al. 2014.

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

Stage	Teacher Activity	Type of items	Number of items	Data Sources		
				DR	O	I
Stages 1 (selecting the assessment target and method) and 2 (implementing ongoing assessment)	Selecting the assessment target	Holistic rubric	1	X		X
	Selecting and implementing the assessment method	Analytic rubric, ratings; final characteristics ratings	22	X	X	X
	Documenting the information collected	Ratings; final characteristics ratings	10	X	X	X
Stage 3 (interpreting data and formulating instructional decisions)	Organizing the data	Checklist, ratings; final characteristics ratings	19	X		X
	Interpreting the data	Holistic rubric, analytic rubric, ratings; final characteristics ratings	9	X		X
	Formulating instructional decisions	Holistic rubric, checklist, analytic rubric	9	X		X
Stage 4 (applying instructional decisions and individualizing)	Applying instructional decisions and individualizing	Holistic rubric	1	X	X	X

Note: DR = document review; O = (video-based) observation; I = 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 the 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 (for example, checklist or rating) that best fit the construct within each stage, choosing the least burdensome level for raters when possible.

Mathematica Ref. No. 40158.B33



Examining Data Informing Teaching (EDIT) Measure Draft Instrument Package

June 2016

Teacher: _ _ _ _ _ _ _ _ Date: _ _ _ _ / _ _ _ _ / _ _ _ _ _ _ _ _	Rater: _ _ _ _ _ _ _ _												
Focal Child 1: _ _ _ _ _ _ _ _ Focal Child 2: _ _ _ _ _ _ _ _	Document review: Start time: _ _ _ _ : _ _ _ _ AM/PM End time: _ _ _ _ : _ _ _ _ AM/PM												
List of artifacts received (Place a "(B)" in front of artifacts if they were collected more than 2 months ago and are background to current performance and progress.)													
<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="border-bottom: 1px solid black; width: 33%;"></td> <td style="border-bottom: 1px solid black; width: 33%;"></td> <td style="border-bottom: 1px solid black; width: 33%;"></td> </tr> <tr> <td style="border-bottom: 1px solid black;"></td> <td style="border-bottom: 1px solid black;"></td> <td style="border-bottom: 1px solid black;"></td> </tr> <tr> <td style="border-bottom: 1px solid black;"></td> <td style="border-bottom: 1px solid black;"></td> <td style="border-bottom: 1px solid black;"></td> </tr> <tr> <td style="border-bottom: 1px solid black;"></td> <td style="border-bottom: 1px solid black;"></td> <td style="border-bottom: 1px solid black;"></td> </tr> </table>													
<small>This collection of information is voluntary and will be used to refine the EDIT measure. Public reporting burden for this collection of information is estimated to average 180 minutes per response, including the time for reviewing instructions, gathering and maintaining the data needed, and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. The OMB control number for this collection is 0970-0355 and it expires 03/31/2018.</small>													

1. HOLISTIC RUBRIC FOR “SELECTING THE ASSESSMENT TARGET”

Assessment targets are linked to meaningful outcomes (that is, a skill, knowledge, or behavior that a child needs to be successful, now or in the future). The targeted behavior is developmentally appropriate (for example, (1) looking at combinations of 5 with manipulatives rather than asking children to solve written equations, or (2) asking children to identify which spoken word is unlike the others in a series of 5 words rather than to compose their own rhyming poem). To meet Head Start and other early childhood education policy requirements, the targets include assessment of key domains related to school-readiness: language and literacy development, cognition and general knowledge, approaches to learning, physical well-being and motor development, and social and emotional development. The targets are generalizable in that the skill, knowledge, or behavior can be demonstrated across settings. The targeted skills or behaviors are those in which children of this age typically make progress within the program year (that is, change is expected in the current year). The targets (1) address areas that are taught in the classroom curriculum, and (2) are defined and measured based on observable behaviors. Targets can be assessed universally (that is, for all children), or evidence of individualization is clear with appropriate accommodations made as necessary for specific children to demonstrate progress.

Sources: Documents (end-of-year goals, assessment objectives or protocols, curriculum/instructional sequence), observations, teacher interview.

1	3	5	7	ENTER RATING
<p>Targets not clearly linked to structured curriculum or meaningful outcomes. OR Targets not developmentally appropriate.</p>	<p>Defines at least one target that is linked to structured curriculum and meaningful outcomes.</p> <p>At least one target is defined for one of the five key school-readiness domains.</p> <p>At least one target is measured using observable behaviors.</p> <p>Targets may not be in an area in which children typically make progress within the program year (that is, cannot measure change over time).</p>	<p>Defines a few targets that are linked to structured curriculum meaningful outcomes.</p> <p>A few targets are individually appropriate for the child.</p> <p>At least one target is defined for each of the five key domains for school-readiness per child.</p> <p>Targets are measured using observable behaviors.</p> <p>At least one target per child is measured over time to track progress.</p> <p>Targets are in an area in which children typically make progress within the program year <i>with instruction or intervention</i> (that is, can measure change over time).</p> <p>At least one assessment target is represented in individualized plans for instruction for at least some of the children.</p>	<p>Defines multiple targets that are linked to structured curriculum and meaningful outcomes.</p> <p>Multiple targets are individually appropriate for the child.</p> <p>Multiple targets are defined for three of the five key domains for school-readiness per child.</p> <p>Targets are measured using observable behaviors AND, as appropriate, teacher examines whether targets generalize across settings.</p> <p>At least one assessment target is represented in individualized plans for instruction for all children.</p>	<p>RATING: _____</p>

Note: For holistic rubrics, even-numbered ratings should be assigned to teachers who exceed the criteria for the preceding rating but do not yet meet all criteria for the subsequent rating. Unless a change or increase is specified, a given rating on a holistic rubric encompasses all criteria for the previous rating.

2. ANALYTIC RUBRIC AND RATINGS FOR “SELECTING AND IMPLEMENTING THE ASSESSMENT METHOD”

Observation or Assessment Method. Teachers decide how they will gather information about the assessment target (learning objective or behaviors). The method of data collection should focus on observable behaviors, be a fair measure of what children know and can do (including linguistic and cultural appropriateness), and offer the opportunity to easily collect information frequently enough that comparisons across time can be made. Over time, the assessment method should provide information about whether the child generalizes the skills, knowledge, and behaviors and so should provide information from more than one context. The frequency of data collection supports reliable and valid interpretation of child progress.

Sources: Documents (assessment objectives or protocols, assessment manual, assessment schedules and plans, description of assessment), observations, interview.

	1	3	5	7	ENTER RATING
A	No evidence available OR The evidence collected is not aligned with assessment target(s).	The evidence collected is occasionally aligned with the assessment target(s). OR Evidence is very limited but is aligned.	The evidence collected is sometimes aligned with the assessment target(s).	The observed evidence collected is always aligned with the assessment target(s).	RATING: _____
B	Assessments not implemented frequently enough to examine child progress.	<u>Only 1 or 2</u> assessment targets are assessed more than once per child to examine progress.	<u>3 or more</u> assessment targets are assessed more than once per child, and frequency of assessments is usually greater when instruction is more intensive.	Assessment information is collected frequently enough to examine child progress (at least three times per reporting period) or change over time. Frequency of data collection is usually greater when instruction is more intensive.	RATING: _____
C	Assessment tasks do not provide information about different levels of the knowledge, skill, or behavior of interest.	For <u>some</u> targets, the assessment task differentiates between varying levels of quality OR independence over time, e.g., by breaking down the task into steps and/or requiring the child to complete the task in different contexts.	For <u>most</u> targets, the assessment task differentiates between varying levels of quality OR independence over time, e.g., by breaking down the task into steps and/or requiring the child to complete the task in different contexts.	For <u>all</u> targets, the assessment task clearly differentiates between varying levels of quality OR independence over time, e.g., by breaking down the task into steps and/or requiring the child to complete the task in different contexts.	RATING: _____
D	All assessments take time away from instruction and practice (>10 minutes).	Assessments typically take time away from instruction and practice (>10 minutes).	Some assessments occur in the context of instructional activities OR in a way that allows teacher to efficiently obtain information about child's knowledge/skill/behavior in a clear task (such as naming pictures, letters, or shapes).	Assessments typically occur in the context of instructional activities OR in a way that allows teacher to efficiently obtain information about child's knowledge/skill/behavior in a clear task (such as naming pictures, letters, or shapes).	RATING: _____ <input type="checkbox"/> CANNOT RATE

	1	3	5	7	ENTER RATING
NOTE: COMPLETE '2E' FOR CHILDREN WITH SPECIAL LEARNING NEEDS OR DUAL LANGUAGE LEARNERS.					
E	Assessment tasks are not valid for child (in terms of language, culture, temperament, and/or ability); teacher does not make needed accommodations or adaptations for children with special learning needs or Dual Language Learners.	For <u>at least one</u> assessment target, teacher assesses in multiple ways, including making needed accommodations or adaptations.	For <u>some</u> assessment targets, teacher assesses in multiple ways, including making needed accommodations or adaptations.	When appropriate teacher assesses target in multiple ways. Teacher documents all accommodations or adaptations used to ensure that the child understands the task.	RATING: _____ <input type="checkbox"/> CANNOT RATE

FOR VIDEO-BASED ASSESSMENT AND INSTRUCTION ONLY

F	Goals of the observed assessment tasks are not consistently clear even to the observer.	Goals of the observed assessment tasks are clear to observers, but the tasks may not be described to the focal child(ren).	<u>Most</u> of the observed assessment tasks are familiar to the focal child(ren) or are clearly described.	<u>All</u> the observed assessment tasks are familiar to focal child(ren) OR teacher ensures focal child(ren) understands tasks before beginning assessments.	RATING: _____
G	In the observed assessment tasks, teacher typically does not devote sufficient attention to focal child(ren) while implementing observation or assessment.	In the observed assessment tasks, teacher typically is frequently interrupted when assessing the focal child(ren).	In the observed assessment tasks, teacher typically focuses on the focal child(ren) being assessed with only small interruptions.	In the observed assessment tasks, teacher typically devotes sufficient attention to focal child(ren) while implementing observation or assessment.	RATING: _____
H	When using a standard task, does not implement in a standard way and does not note changes in administration.	When using a standard task, sometimes implements in a standard way and sometimes documents the deviations from standard administration, if applicable.	When using a standard task, usually implements in a standard way. Consistently documents any deviations.	When using a standard task, implements in a standard way (with fidelity). Documents if concerned about valid response, if applicable.	RATING: _____ <input type="checkbox"/> NOT APPLICABLE <input type="checkbox"/> CANNOT RATE

NOTE: For analytic rubrics, even-numbered ratings should be assigned to teachers who exceed the criteria for the preceding rating but do not yet meet all criteria for the subsequent rating.

RATINGS: Indicate how characteristic each item is. (Note: check the appropriate category.)

Sources: Documents (assessment objectives or protocols, assessment manual, assessment schedules and plans, description of assessment), observations, interview.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	CANNOT RATE
I	Assessments typically occur in a familiar context.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	0 <input type="checkbox"/>
J	Child's family or household members help collect assessment information.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	0 <input type="checkbox"/>
<u>DOES THE TEACHER INDIVIDUALIZE INSTRUCTION?</u>						
<p>1 <input type="checkbox"/> Yes</p> <p>0 <input type="checkbox"/> No → GO TO N</p>						
K	Teacher monitors child's progress in area of individualization with at least 3 pieces of evidence (data points).	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	
L	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.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	
M	Teacher continues to periodically monitor child's progress in area of individualization even after first signs of improved progress.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	
<u>FOR VIDEO-BASED ASSESSMENT AND INSTRUCTION ONLY</u>						
N	Teacher documents child behavior/performance or collects work sample.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	0 <input type="checkbox"/>

3. RATINGS FOR “DOCUMENTING THE INFORMATION COLLECTED”

When teachers implement ongoing assessments, they need to document child progress objectively, accurately, efficiently, and with relevant contextual information.

Sources: Documents (portfolios, checklists, anecdotal records), observations, interview.

RATINGS FOR FOCAL CHILDREN

Indicate how characteristic each item is. (Note: check the appropriate category.) If this differs by child please note separate ratings for each children.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
A	Documentation provides relevant information about the general assessment context, supports, and timing.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
B	Documentation includes child-specific context, including types of prompts and supports.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
C	The documentation can be understood without the need to ask the teacher questions beyond clarifying shorthand codes.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

RATINGS: Indicate how characteristic each item is. (Note: check the appropriate category.)

		FOCAL CHILD 1				FOCAL CHILD 2			
		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
D	Documentation includes information about the child's strengths.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
E	Documentation includes errors and misconceptions.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
F	Documentation is objective, describing what happened rather than making subjective comments and judgments.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
G	Documentation is an intentional process, with method selected when planning instruction.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

DOES THE TEACHER HAVE INDIVIDUAL LEARNING PLANS?

1 Yes
 1 No → **GO TO NEXT SECTION**

H	Documentation includes child's responses/performance during individual learning plan activities.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
----------	--	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------	----------------------------

4. CHECKLIST AND RATINGS FOR “ORGANIZING THE INFORMATION COLLECTED”

The teacher’s organization of the documentation should facilitate interpretation and communication with families and team members, impose a minimal burden on teachers, and provide consistent, reliable data entry.

Sources: Documents (documentation from assessment, data display), teacher interview.

CHECKLIST

		YES	NO
A	The teacher views/organizes the data to compare a child’s performance to a developmental expectation or benchmark for growth.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
B	The teacher views/organizes the data to compare a child’s progress to a developmental expectation or benchmark for growth.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
C	The teacher views/organizes the data to compare a child’s performance to that of other children in the class.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
D	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.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
E	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).	1 <input type="checkbox"/>	0 <input type="checkbox"/>
F	The system for organization is efficient. (For computer systems, answer will typically be “yes” if teacher is easily able to access and enter information. Note computer application or describe below.)	1 <input type="checkbox"/>	0 <input type="checkbox"/>
G	Teacher presents/organizes the data to communicate to parents about the child’s strengths and weaknesses.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
H	The teacher organizes the information to look at performance by class for one or more assessment targets at a single timepoint.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
I	The teacher organizes the information to look at progress by class for one or more assessment targets across multiple timepoints.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
J	The teacher organizes the information to look at performance by subgroup for one or more assessment targets at a single timepoint.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
K	The teacher organizes the information to look at progress by subgroup for one or more assessment targets across multiple timepoints.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
L	The teacher organizes the information to look at performance by domain for the class or a subgroup for one or more assessment targets at a single timepoint.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
M	The teacher organizes the information to look at a child’s progress by domain within reporting periods (multiple assessment targets across multiple timepoints).	1 <input type="checkbox"/>	0 <input type="checkbox"/>
N	Teacher uses visual displays to depict child progress by date for at least one target.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
O	Teacher indicates in documentation when a new instructional strategy or individual learning plan is implemented.	1 <input type="checkbox"/>	0 <input type="checkbox"/>

RATINGS

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	NOT APPLICABLE
P	The teacher views/organizes data in a way that makes it easy to understand the current level of the child's skill compared to prior checkpoints/reporting periods.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	
Q	If the assessment system organizes the data, the teacher uses the system with fidelity—for example, the teacher enters information within a week. <i>Make notes on the time frame.</i>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	0 <input type="checkbox"/>
R	If using a teacher-developed system, the teacher files or enters data on at least a weekly basis.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	0 <input type="checkbox"/>

5. HOLISTIC AND ANALYTIC RUBRICS AND RATINGS FOR “INTERPRETING THE DATA”

Teachers draw on multiple sources of information in making decisions about the current skills, knowledge, or behavior of children and the progress they are making. The teachers use objective, reliable data and consider alternative explanations for children’s behavior. Teachers consider the pattern of data and set hypotheses about children’s development and instructional needs.

Source: Documents, teacher interview (with questions about the video).

	1	3	5	7	ENTER RATING
A	Bases ratings on impressions or memory. No evidence to support interpretation.	Bases most ratings on single piece of information (that is, taken at a point in time or single event). Evidence is too scanty to support interpretation. OR Some evidence supports interpretation whereas other evidence does not.	Bases most ratings/ inferences on at least two pieces of information (can be from a single context or task). Evidence supports interpretation for most assessment targets, but evidence for some interpretations is based on inference rather than objective data. Considers alternative explanations of observed skill/behavior, including the amount of support available to the child. Considers child’s performance and progress relative to typical development or progress of peers. Examines more than one data point in between checkpoints for at least one target per child (such as preliminary ratings).	Bases most ratings/ inferences on multiple pieces of information collected in more than one context or task. Evidence is objective and supports interpretation. Considers context of observed skill/behavior. Considers child’s strengths and interests as well as challenges. Tests some alternative explanations of observed skill/behavior. Considers child’s performance and progress relative to typical development and progress of peers.	RATING: _____
B	Infrequent examination of data (does not examine child’s records and make ratings at reporting periods).	Examines current data in relation to performance criteria, but does not consider the child’s rate of progress.	Evaluates progress (that is, change in performance within reporting periods). Uses data to identify areas of learning where progress is slower than expected and where it is faster than expected.	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.	RATING: _____

RATINGS

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
C	Teacher involves the family in interpreting and understanding the data.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
D	Teacher involves the other teachers/staff in interpreting and understanding the data.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

6. HOLISTIC RUBRIC, CHECKLIST, AND ANALYTIC RUBRIC FOR “FORMULATING INSTRUCTIONAL DECISIONS”

Based on the interpretation of the data and any hypothesis set, teachers plan instructional activities and strategies to support increased or continued growth. They plan instructional groups intentionally to support children’s development (e.g., partnering children to provide a good language model for a child who needs to talk more, targeted instruction of small groups of children with a similar need). Instruction considers what data suggest about children’s current developmental status and progress by, for example, providing more independent practice in areas where a child is successful, testing hypotheses about the types of support children need, or considering whether children generalize skills across tasks.

Sources: Documents (data display, assessment records, lesson plans), teacher interview (with questions about the video).

RUBRIC

	1	3	5	7	ENTER RATING
A	Instruction may be designed to be interesting and engaging, but it does not draw on what is learned from the data or follow a logical plan for developing child(ren)’s skills.	<p>Instructional decisions follow curriculum recommendations.</p> <p>Instruction includes more than one small group or individual lesson/planned interaction.</p> <p>Instructional grouping is based mainly on social and behavioral indicators (such as friendships or avoiding problem pairs).</p>	<p>Instructional groups and activities are designed intentionally using data to support child(ren)’s development.</p> <p>Instruction follows a logical sequence to increase child(ren)’s skills across time.</p> <p>Instructional decisions consider the evidence of current developmental status for each individual child.</p> <p>Some instructional decisions draw on evidence-based or professionally recommended strategies.</p> <p>Instructional decisions provide more intensive instruction in areas where child(ren) lag behind developmental expectations for growth.</p>	<p>Instructional decisions include modifications and adaptations based on the evidence of child development including patterns of progress across time and consideration of alternative explanations for child performance.</p> <p>Instructional decisions consider child(ren)’s interests and strengths as well as areas for growth.</p> <p>Considers child’s progress relative to typical development and progress of peers.</p> <p>Reaches out to external resources as needed.</p> <p>Most or all instructional decisions draw on evidence-based or professionally recommended strategies.</p>	RATING: _____

CHECKLIST

Is there evidence of varied approaches to individualization that are responsive to the data?

SELECT ALL THAT APPLY		DOES NOT OCCUR	OCCURS AND NOT DOCUMENTED	OCCURS AND DOCUMENTED	CANNOT RATE
B	Prompting or questioning strategies	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
C	Additional practice	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
D	Grouping strategies (homogeneity for a specific learning need)	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
E	Peer supports (heterogeneity)	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
F	Adult supports	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
G	Environmental supports (e.g., physical positioning of child)	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
H	Other adaptations or modifications designed to meet specific child needs (<i>Specify</i>): _____	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>

RUBRIC

Rate overall the teacher's use of strategies.

	1	3	5	7	ENTER RATING
I	No evidence in the data to suggest a need for the individualization.	Some evidence suggesting a need, but no evidence suggesting that it is supporting the child's progress and teacher did not try different strategies (allow a 2 week trial period to look for change the strategy before trying something different). OR Some evidence suggesting a need but no individualization implemented.	Evidence that the child needed the support and is either making progress or the teacher is continuing to try different strategies.	Evidence that the teacher is using the data to make decisions about the success of strategies and changes those that are not effective in supporting the child.	RATING: _____

7. HOLISTIC RUBRIC FOR “APPLYING INSTRUCTIONAL DECISIONS AND INDIVIDUALIZING”

The approach is responsive to the data (and progress) and the interpretation of the data. The teacher collects further evidence to evaluate whether the instructional approach is valid for meeting the targeted instructional need for this child (or these children). That is, the teacher notes instructional changes and assesses progress to evaluate the success of those changes. The teacher may use flexible evidence-based (or professionally recommended) strategies, such as using purposeful small groups, offering different questions or different levels of prompts, offering more practice for a child in a particular area throughout the day, adapting and modifying the activity, and using different instructional approaches (such as varied levels of visual or auditory cues or individual-versus-group interaction). The teacher may provide peer, adult, and environmental supports. Whenever possible, the teacher incorporates and builds on children’s strengths and interests when individualizing.

Sources: Documents (lesson plans, plans for individualized instruction), observations, interview.

1	3	5	7	ENTER RATING
<p>Provides standard instruction for every child—that is, does not make any changes for any child (no evidence of individualization or differentiation).</p>	<p>Increases opportunities for practice for children with weaknesses identified in the data.</p> <p>Shows evidence of an intentional adaptation for at least one child.</p> <p>Uses as least one evidence-based or professionally recommended practice.</p> <p>Collects at least one piece of data about a child’s response to individualized goals and instruction.</p>	<p>Increases or varies opportunities for practice for some children with weaknesses identified in the data.</p> <p>Uses some varied instructional strategies that are evidence-based or professionally recommended.</p> <p>Provides support for emerging skills identified in the data, recognizing when children need a challenge.</p> <p>Changes instruction if child is not demonstrating improvement or greater progress.</p> <p>Classroom instructional team has a shared knowledge about goals and instructional strategies for each child.</p> <p>Plans and collects at least two pieces of evidence about how each child responds to individualized goals and instruction.</p>	<p>Increases or varies opportunities for practice or level of scaffolding for most children with weaknesses identified in the data.</p> <p>Uses varied instructional strategies that are evidence-based or professionally recommended to build on strengths and mitigate weaknesses.</p> <p>Incorporates child interests and experiences.</p> <p>Plans and collects at least three pieces of evidence about how each child responds to individualized goals and instruction.</p> <p>Organizes and reflects 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).</p>	<p>RATING: _____</p>

8. CHARACTERISTICS RATINGS

GENERAL GUIDELINES FOR RATING CHARACTERISTICS

Rate how characteristic the statement is of this teacher/classroom.

1	Not At All (almost never)	You did not see this at all, or it was seen only once or twice and it is a behavior/type of documentation that usually occurs frequently in classrooms.
2	Minimally characteristic (sometimes evident)	Rate here if you see it occasionally and it is appropriate at many other times, or you see it happen sometimes, and it is only partially present (for example, some information about context on a few pieces of documentation but never complete documentation) or present only for a single assessment target/learning objective.
3	Strongly characteristic (frequently evident)	Rate as 3 if something happens frequently and across domains (as appropriate), but does not occur at all the appropriate times. Also rate here for something that would normally be 'low frequency' but is evident at many of the appropriate times or in most of the appropriate documents.
4	Extremely characteristic (almost always evident)	Something that happens at appropriate times and in appropriate documents across multiple domains. If something happens frequently but is missing in some appropriate documents or observations, then code as 'strongly characteristic' rather than 'extremely characteristic'.

SELECT ONE PER ROW

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
A	Teacher collects information that allows him/her to see whether child's current rate of progress is higher than the child's rate of progress prior to a change in instruction.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
B	The teacher uses data to examine differences in rates of progress relative to major changes in instruction (organizes it and interprets the change that does or does not occur relative to changes in instruction).	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
C	The teacher involves the family in helping to collect assessment information.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
D	The teacher collects assessment data (that is, documentation of child behavior/performance) in an efficient manner so that s/he minimizes time away from interacting with children.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
E	The documentation includes description of errors, misconceptions, or early steps towards a learning goal.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
F	The teacher examines the data in different ways (by child across domains, by domain or learning objectives across subgroups and children) to understand the child's performance in different ways.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
G	The teacher is able to articulate both the strengths and weaknesses of individual children.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
H	The teacher can support inferences about strengths and weaknesses <i>with evidence from the assessment data</i> .	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

SELECT ONE PER ROW

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
I	Documentation is objective, describing actual behaviors rather than subjective comments.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
J	The teacher clearly communicates both strengths and challenges of the child with the family, providing evidence for each of these.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
K	The teacher involves the family in interpreting the data across multiple domains. ^a	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

DOES THE TEACHER HAVE A PLAN FOR DATA COLLECTION?

1 Yes
 0 No → **GO TO P**

L	The teacher's plan for data collection assures that all children are assessed periodically within each reporting period across domains.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
M	The teacher follows the plan/schedule for collecting assessment data each month that assures data is collected on each child across more than one domain.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
N	The teacher's data collection plan identifies certain instructional targets on which specific children get more frequent assessment (areas of individualization).	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

IS THERE EVIDENCE OF INDIVIDUALIZATION?

1 Yes
 0 No → **END**

O	The teacher monitors child progress in their area of individualization with at least 3 pieces of evidence for each reporting period (approximately each quarter).	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
P	Every child is assessed in at least one area of individualization.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

^aFor item K, rate 4, if evidence of family involvement is found in multiple domains and the teacher invites family's interpretation and documentation; 3, if at least one domain or objective involves family as partner [rather than recipient of info] 2, if teacher provided interpretation of data to family [family as recipient of info]; and 1, not at all [information for families does not explain what led to the interpretation; or there is no communication with family].

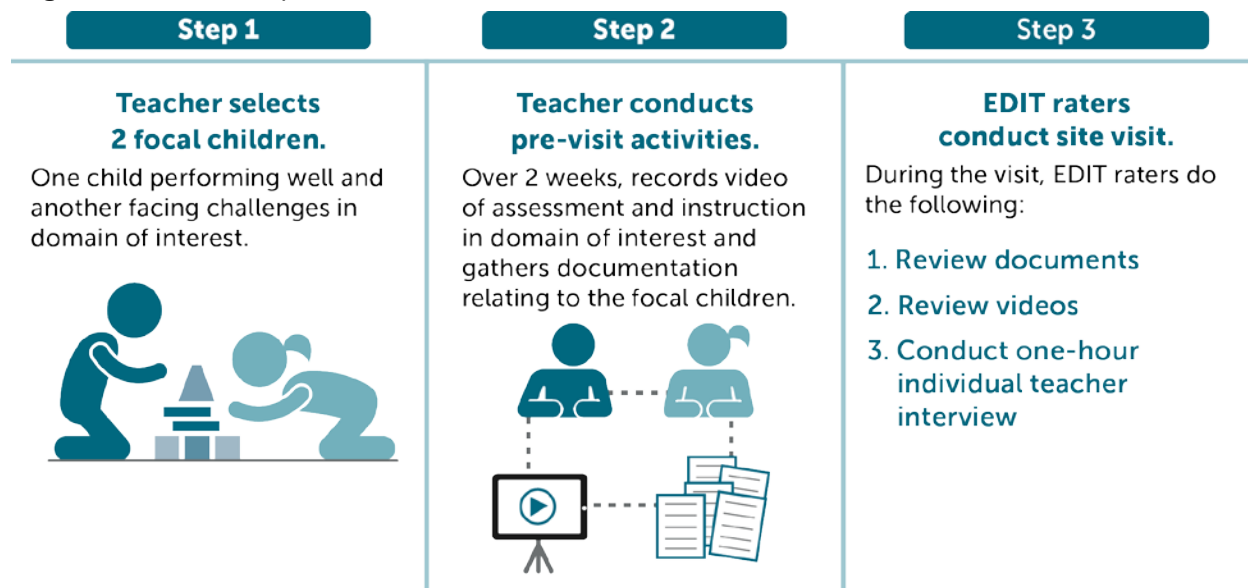
An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. The OMB control number for this collection is 0970-0355 and it expires 03/31/2018.

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

D. Administration procedures

The EDIT is designed for use in preschool classrooms using a curriculum-embedded approach to ongoing child assessment. The teacher and children need at least two months in the classroom before administering the EDIT in order to have evidence of progress monitoring. In many programs, the first month is spent screening and getting to know the children; ongoing assessment starts after the first month. In this section, we discuss each step in the administration of the EDIT (Figure II.2).

Figure II.2. EDIT procedures



1. Step 1: Selecting focal children

The team asks each teacher to select two focal children from consented children in the classroom. The teacher is asked to select one child who is doing well and one child who has challenges in language or literacy. Both children need to have been in the classroom for at least two months.

2. Step 2: Teacher activities before the raters' visit

Assembling documents. Lead teachers in each classroom receive a letter explaining the study activities. The letter includes a description of the documents that the team would like the teachers to collect before the raters' visit to the program (see Appendix C 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 documentation on children performing well and the children facing challenges when teachers selected focal children versus when the team used stratified random selection of children.¹⁴

Learning to use the video equipment. Teachers are shipped an electronic tablet and accessories (including a charger and desktop 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 during the two weeks before the raters' site visit. The team asks that the teacher video record the following sequence (Figure II.3) for one child performing well and another child facing challenges, resulting in six data points collected across two weeks.¹⁵

Figure II.3. Method for the focus and frequency of video recordings

	Week 1		Week 2	
Child Performing Well	Language/Literacy or Social Skills Instruction (Individual or Small-Group)	Assessment	Language/Literacy or Social Skills Instruction (Individual or Small-Group)	Assessment
Child Facing Challenges	Language/Literacy or Social Skills Instruction (Individual or Small-Group)	Assessment	Language/Literacy or Social Skills Instruction (Individual or Small-Group)	Assessment

For the instruction activities, the team asks that language, literacy, or social skills domains 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. Step 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

¹⁴ In the initial rounds, the team asked teachers to complete a language and literacy screener for all children and stratified by their scores. Our plan was to randomly select from the upper and lower quartiles, however, there was limited variance on the screener with most children scoring highly.

¹⁵ If the teacher would normally group the two focal children for small-group instruction, the teacher may use the same small-group instruction videos for both children.

- 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

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.

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 about 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.4). 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 videos 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.4. Use of data sources to assign scores

First, raters review documents and assign scores.	Second, raters review the videos and adjust scores.	Third, raters conduct teacher interview and finalize scores.
<p>After reviewing the documents, EDIT raters assign preliminary scores where possible. Raters note missing information to look for in the observations and probes to include in the teacher interview.</p> 	<p>After watching the video-recorded observations, raters adjust or assign scores as needed. Raters note additional probes for the teacher interview.</p> 	<p>After the interview, raters finalize all scores.</p> 

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 team for further discussion.

Interviewing the teacher. The EDIT rater conducts a one-hour 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 raters' 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 or at the end of the day after the children are picked up. Conducting the interview during children's naptime sometimes is feasible when classroom coverage is available. 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. The interview is conducted in person so documents can be shared as needed. After the interview, the EDIT rater spends approximately an hour finalizing scores on all items.

CHAPTER 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 six rounds of data collection conducted with 18 English-speaking early childhood teachers in 13 centers with video recorded observations of the teachers working with the children in their classrooms. Eleven of the classrooms were Head Start and the remaining seven were a mix of state-funded and private preschool classrooms¹⁶ (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
Data collection round 4: December 2015–January 2016	
Center G	Classroom 10
	Classroom 11
Center H	Classroom 12
Data collection round 5: February 2016	
Center I	Classroom 13
	Classroom 14
Center J	Classroom 15
Data collection round 6: April–May 2016	
Center K	Classroom 16
Center L	Classroom 17
Center M	Classroom 18

¹⁶ Two of the lead teachers were Spanish-English bilingual, and seven of the classrooms had children from households with Spanish speakers.

The EDIT study team sought centers using a curriculum-embedded ongoing assessment system with moderate to high quality implementation. The team chose the centers purposively based on recommendations from research colleagues who had been in classrooms at the centers. After analysis of our first nine pretest visits, we increased the number of EDIT items and indicators that measured teacher planning and intentionality when conducting assessment and individualizing instruction. To test those items in the remaining visits we prioritized classrooms likely to exhibit those practices, including classrooms that administered both curriculum-embedded assessment and GOMs¹⁷, when available. In general, the researchers thought the centers we visited were implementing ongoing assessment with moderate to high fidelity. The centers were in six states.

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 study team refined 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. By round 3, the team settled on a standard set of procedures that appeared useful.

1. Recruitment and consent

The study team began all recruitment efforts by sending a letter in advance 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 local Setting Point Person (SPP). After recruiting the setting, the team spoke with the SPP as needed to check on consent status, that teachers received materials, and to schedule and confirm visit dates. The SPP also identified teachers to participate in the EDIT; criteria for selection within each center varied, because it was at the discretion of the setting contact. Next, the team shipped consent packets—including Institutional Review Board (IRB)-approved consent forms¹⁸—to the SPP to distribute to the participating lead teachers, who, in turn, distributed consent packets to parents of the children in their classrooms. 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 of 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 preselected 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 teachers sometimes video recorded

¹⁷ GOMs are general outcome measures. Please see page 5 for additional description.

¹⁸ Consent was sought from teachers and from parents for permission for the video-based observations and to share documents about the focal children. During round 1, consent was also sought from parents for the teacher to complete a language and literacy screening measure for all children. The study also received a generic clearance from the U. S. Office of Management and Budget for the final 9 visits.

outside the classroom because it was difficult to ensure that only consented children were visible in the video. In rounds 3 through 6 of the pretest, the team once again sought consent from all children in the classroom. This allowed teachers to video record more naturally occurring assessment and instruction even though it took more time to collect consent.

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

Data collection activities	Round 1	Round 2	Rounds 3–6 ^c
EDIT advance letter to setting contact.	Δ	Δ	Δ
Study team calls setting contact and identifies SPP. ^a	Δ	Δ	Δ
If SPP is different from initial setting contact, study team calls SPP.	Δ	Δ	Δ
Study team sends consent packets to SPP, and SPP distributes consent packets to teachers.	Δ		Δ
Teachers/SPP collect consents and ship them to study team.	Δ		Δ
Study team sends OL&C ^b packets to teachers for consented children.	Δ		
Teachers complete OL&Cs and send results to study team.	Δ		
Study team randomly selects two focal children stratified by performance on OL&C.	Δ		
Study team ships teachers video recording materials with identification of focal children and schedule for video recording.	Δ		
Study team ships teachers video recording and consent materials and a schedule with instructions to select two focal children and return consent for the teacher and focal children before beginning additional activities.		Δ	
Study 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
Study team offers telephone support for teachers.	Δ	Δ	
Study team requests call with teacher to discuss video recording.			Δ
Teachers conduct video recording. ¹⁹	Δ	Δ	Δ
Teachers gather documents.	Δ	Δ	Δ
Study team conducts site visit with two raters.	▲	▲	▲
Study team holds debrief calls with teachers.	□	□	□

^aSetting Point Person (SPP) is a primary point of contact at each center.

^bOral Language & Comprehension Questionnaire (adapted from Bradfield and McConnell 2013).

Δ Before visit ▲ During visit □ Post-visit

^cRounds 3–6 are combined because the procedures did not change. The procedures implemented in rounds 3–6 represent what we learned from rounds 1 and 2.

¹⁹ Teachers were instructed to only include consented children in their video recorded observations.

Table III.3. Consent collection during pretest

Data collection round	Consent from lead teacher	Consent from all children in classroom	Consent from 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	
Data collection round 4: December 2015–January 2016	X	X	
Data collection round 5: February 2016	X	X	
Data collection round 6: April–May 2016	X	X	

2. Selecting focal children

The study team pretested a method for selecting two focal children (one performing well²⁰ 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 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. The study 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.²¹ However, the OL&C was intended for a different purpose: identifying children who are struggling, but not necessarily ranking children who are performing well.

In the remaining five rounds of data collection, teachers were not asked to complete the OL&C, and the study 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. In addition, many children received high ratings on the OL&C screener making it difficult to randomly select children with very different skill levels. We found greater differences in the teacher selected children's skills.

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 before the team's visit to the program. The letter asked teachers to gather any documentation they had that 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

²⁰ By "performing well," we refer to children meeting or exceeding developmental expectations for their age.

²¹ 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.

(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 desktop 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.²² 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 whether the video was of assessment or instruction, timing, and number of videos per child. In the first round of data collection, the team examined two video recording methods that varied in the instructional or assessment focus and frequency of the video recordings (Figure III.1). Each teacher was assigned to only one method.

²² The team also asked the teachers to check that the video recording was successful (that is, that the recording was not blank).

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

		Week 1	Week 2	Week 3		
Method 1 (n = 16 teachers) Used in rounds 1-6	Child Performing Well (n = 16 children)	Language/ Literacy or Social Skills Instruction (Individual or Small-Group)	Assessment	Language/ Literacy or Social Skills Instruction (Individual or Small-Group)		
	Child Facing Challenges (n = 16 children)	Language/ Literacy or Social Skills Instruction (Individual or Small-Group)	Assessment	Language/ Literacy or Social Skills Instruction (Individual or Small-Group)		
Method 2 (n = 2 teachers) Used in round 1	Child Facing Challenges (n = 2 children)	Language/ Literacy or Social Skills Instruction (Individual or Small-Group)	Assessment	Language/ Literacy or Social Skills Instruction (Individual or Small-Group)	Assessment	Language/ Literacy or Social Skills Instruction (Individual or Small-Group)

- **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,²³ 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 predicted that changes made for low-performing children would be more obvious to the raters. 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.

²³ 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 the remaining five rounds 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 sequence of videos than requested.

4. Raters' visit

During the pretest, two raters from the EDIT study team 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 semistructured interview with the lead teacher. The EDIT study team consisted of eight researchers. Seven of the researchers paired in different combinations for each visit. 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 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. Chapters IV and V summarize information about participants who assisted with development of the EDIT.

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 and gathering documents, as well as ease of video recording.

C. Refining the instrument and its procedures

After each round of data collection, the EDIT study 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. Chapter V contains a description of the revisions.

In between the second and third round of data collection, the EDIT study team shared the EDIT with the project's expert panel for its 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. Additional, feedback from the expert panel was also incorporated at the conclusion of data collection.

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

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

A. Sample: Who participated in the pretest?

As noted in Chapter III, we conducted the pretest with 18 purposively selected²⁴ classrooms, visiting 8 classrooms in spring 2014, one classroom in fall 2014, and 9 classrooms in spring 2016. We found that the EDIT could be used in the fall and spring of the program year to identify areas for growth and capture strengths. The classrooms were 11 Head Start classrooms (from seven Head Start centers) and 7 preschool classrooms in private programs in six states. Of the classrooms we visited, 2 were classrooms of 3-year-olds, 5 were classrooms of 4-year-olds, and 13 were classrooms of mixed ages ranging with most children being 3 to 4.5-years old.²⁵ Seven 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 an SPP in each setting who answered questions about the children in the classroom (for example, age ranges and 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 SAQ. (Appendix B 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 18 teachers, 17 worked full-time. All lead teachers were female, ranging in age from 30 to 57. Two teachers were Spanish-English bilingual. All the teachers had an associate's college degree or higher. Of the 18 teachers, 14 (78 percent) had at least a bachelor's degree, and 8 of those teachers also had master's degrees. This was a highly educated group compared to, for example, data from the National Survey of Early Care and Education, which indicated that 17 percent of teachers for children 3 to 5 years old had an associate's degree and 45 percent had a bachelor's degree or higher (National Survey of Early Care and Education 2013). The teachers in our pretest had early childhood teaching experience ranging from 2 to 32 years, with a median of 5 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 one to 40 hours a year.

²⁴ Settings were selected based on their use of ongoing assessment systems.

²⁵ Mixed-aged classrooms ranged from 2 to 6-year-olds.

Sixteen teachers also reported having mentors for these topics, and most teachers reported meeting with their supervisor about these topics monthly to 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. Of the 18 teachers, 13 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

All classrooms used ongoing assessment systems. Of the 18 classrooms, 10 used Teaching Strategies GOLD (TS GOLD; Teaching Strategies, Inc. 2011). Six classrooms used the Work Sampling System (WSS; Meisels et al. 2001).²⁶ One classroom used Tools of the Mind (Bodrova and Leong 2007). One classroom used a school-developed system for ongoing assessment. Despite differences in TS GOLD and WSS (Table IV.1), we captured similar information on the EDIT across both systems. The most recent Head Start Program Information Report (PIR) data (2014–2015) indicated that TS GOLD was the most frequently used preschool ongoing assessment system in Head Start nationally. According to the 2014–2015 PIR data, more than half of all Head Start programs report using the preschool TS GOLD or its precursor, the Creative Curriculum Developmental Checklist. WSS is used less frequently, with 2 percent of Head Start programs reporting using it.

²⁶ 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 TS GOLD or WSS.

Table IV.1. Two ongoing assessment systems

Characteristic	Ongoing assessment system	
	Teaching Strategies GOLD	Work Sampling System
Areas measured by the assessment system	Social-Emotional Language Literacy Cognitive Math Science and Technology Social Studies Arts Physical Each of the areas is composed of objectives and dimensions.	Personal and social development Language and literacy Mathematical thinking Scientific thinking Social studies Arts Physical development 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. ²⁷
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 kindergarten. Scores are levels from 1 to 9. 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." Children can be scored using criteria for an earlier age range to meet developmental needs.
Age/grade level	Birth to age 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 received ratings above the midpoint 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 be enough for some items. One teacher we visited scored low on most items,

²⁷ Our final pretest visit indicated that the most recent version of WSS (fifth edition) no longer includes a separate portfolio.

allowing us to fine-tune the low end of the scales. Future users should confirm that the low end indicators adequately discriminate among the initial levels of quality.

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, during the first nine pretest visits, most scores were on the high end of the scales. During the final nine pretest visits, we achieved greater item variance. A combination of modifications to the items along with a more diverse sample helped us achieve this greater variance as the pretest progressed. For example, item 7 was measured on a 7-point scale. The mean score on item 7 for the first nine pretest visits was 5.29. After revising this item to include more stringent criteria about the teacher assessing the impact of individualization, the mean score was 4.3 (closer to the midpoint of the scale) for the final nine pretest visits.

1. Items with high mean scores

In this sample of teachers from recommended programs, we identified items that measure teacher practices that teachers are more likely to do with quality. Our examination of assessment documents and observations indicates high scores on general assessment practices that are important for summative and performance-based assessment and that also can inform monitoring of progress. During our first nine pretest visits, the most positively rated practices involved teachers' documentation and organization. After changes to the EDIT to assess intentionality and provide greater description of the high quality criteria, the stringency of the criteria increased. However, during the last nine visits, documentation remained among the most positively rated assessment practices that teachers implemented.

Teachers' documentation. Throughout the pretest, documentation remained among the most positive practice based on EDIT scores. 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, but several teachers also had other documentation. One teacher did not have an electronic assessment system. Teachers entered documentation ranging from daily, to weekly, to whenever 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. With a broad sample of teachers, however, we would expect that this is an item that will have strong variance in scores.

Teachers' organization. During the first nine visits, organization was among the most positive practices based on EDIT scores.²⁸ Both TS GOLD and 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 TS GOLD, the displays of data included criteria for expected levels of performance at different ages. 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.²⁹ 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. Throughout the pretest, there were aspects of instruction and individualization that remained among the most positive behaviors based on EDIT scores. Teachers wrote individual instructional 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 (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. With a few exceptions, however, we did not hear evidence of intentionally selecting these strategies based on the information the teacher had about children's progress. We found that the EDIT did not reflect this difference in intentional use of data well. During the first nine visits, teachers received high scores on many of the items assessing the quality of data collection and organization of data, and plans for individualization even when intentionality was not evident. After this, we revised a number of EDIT items to clarify criteria and increase measurement of intentionality (described in depth in Chapter V) and found greater variance in performance.

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

In reflecting on what we observed among teachers during the first three rounds of data collection, 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.

²⁸ After these visits, we made revisions to EDIT items (described in Chapter V). These revisions, along with more diverse sample, are likely factors in the greater variance we saw in these items during our final nine visits.

²⁹ The two teachers who did not use TS GOLD or WSS did not have data displays but instead used their professional judgement to inform their selection of goals for the children.

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. We expanded our measurement of intentionality for our final nine visits (see Chapter V).

Before making our initial 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. We subsequently expanded our measurement of teacher planning.

Most 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 in a portfolio. 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. 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 rarely observed teacher-developed checklists that were used on a regular basis. The limited use of any type 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 much planning for data collection. The two teachers who developed their own checklists were highly experienced teachers, each with at least 15 years of experience teaching in early childhood settings.

Teachers' instruction and individualization. When planning what skills and knowledge to address in individualizing, teachers most often 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. Only two of the teachers articulated expectations for progress within a single reporting period. Instead, most of the teachers looked at the 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 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." We did meet one teacher who specifically had a rule about always collecting three pieces of evidence per area of individualization. We met another teacher who continued to collect data on each area of individualization until the teacher was able to consistently document the child's mastery in multiple contexts.

Teacher awareness of how to identify evidence-based or professionally recommended teaching strategies to individualize instruction was limited. We probed this area to make sure that what we observed was capturing teacher practice. Fourteen 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 or from training in a particular approach (for example, Montessori training). This suggests that the EDIT may provide important information about areas in which staff need more professional development or support for implementing practices.

After completing the first two rounds' pretest visits, 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 to the EDIT items asking if the supports and prompts were documented and if they seemed responsive to the data. In the first three rounds of pretest visits (rounds 1–3)—our initial 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 we 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. During our final three rounds of pretest visits (rounds 4–6), we did meet at least two teachers who were intentionally continuing the loop back to stage 1, and the EDIT was able to measure and distinguish these behaviors among teachers.

Exhibit IV.1. Study team reflections on some pretest teachers’ use of assessment

Many 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

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

- These teachers used TS GOLD, the WSS ongoing assessment, Tools of the Mind, or a school-developed system as part of their ongoing assessment practice.
- The teachers' use of these systems was largely supported through training and mentors.
- Teachers had higher levels of formal education than U.S. early childhood educators, in general.

Overall, the initial teachers in rounds 1–3 received ratings above the mid-point on many 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. We revised the EDIT accordingly in rounds 4–6 and found more variation in ratings.

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 often limited to one or two observational records, and the method for collecting these observational records often was inefficient and did not require planning.
- In many cases, teacher awareness of how to identify evidence-based or professionally recommended strategies was limited to use of curricular materials and colleagues.
- Teachers rarely articulated expectations for child progress (versus performance) within a single reporting period.

Evidence of planning and intentionality in collecting and using data to inform instruction and individualization was usually 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.

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

CHAPTER V. LESSONS LEARNED FROM THE PRETEST

The goal of pretesting 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 study team by telephone. Of the 18 teachers, 13 participated.³⁰ The debrief discussions gave teachers an opportunity to comment on the research participation process, including the burden of collecting consents and gathering documents, as well as ease of video recording.

All who participated in a debrief call reported they would recommend participating in the study to colleagues. Some teachers reported that the project looked overwhelming when they first received our materials, but they reported minimal burden after they got started. We have since streamlined our materials (for example, reduced the text and added graphics) to attempt to portray what is involved without overwhelming the teachers (Appendix C). 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 one to eight videos ranging from one to 18 minutes, with an average total recorded length of 31 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 the interviews taking a full hour. These interviews typically took place late in the school day (for example, during naptime or at the end of the 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	1 hour
Recording videos	1 hour
Teacher interviews	1 hour
Total time	3 hours

³⁰ All 18 teachers were invited to participate in a debrief conversation; however, we only conducted interviews with teachers who could speak to us within our data collection time frame.

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 study team select two children.	Teachers self-selected 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" and "instruction" documentation.	Teachers were instructed to gather "things that you collect and use to help decide how to support the learning of the ...children including any observations, assessments, and plans for instruction."
Time period for documentation	Raters use worksheets to count data related to assessment targets in documentation gathered by teacher.	Emphasized the previous 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, alternating the order of instruction and assessment activities	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

After the first pretest round, 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 2015. 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 two 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 10 pretest visits, 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, because this appeared to capture greater differences in child performance.

2. Describing and accessing documentation

Initially in our first round, 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. Based on the way 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 two teachers 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, it is beneficial to attempt to obtain permission for all children in a classroom or provide a means for teachers to obscure the names of other children (although this would increase teacher burden).

It is also beneficial to interview the teacher in her classroom when possible. Interviews conducted in the classroom yielded richer information. In six of our visits, we were able to 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, portfolios, 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 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.

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. In one case, the teacher asked a colleague for assistance in accessing data during our visit.

An additional consideration was access to the version of the classroom's ongoing assessment rubrics. TS GOLD is in its first edition and 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. In the classrooms using the most recent version of WSS, it was difficult to evaluate whether the anecdotal records provided evidence aligned with the indicators. We recommend asking programs to provide raters with 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 collection³¹) 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. For example, the same anecdote might be entered for multiple objectives, multiple children, and sometimes for an entire classroom (based on a group activity). The counts of anecdotal records could overestimate the evidence available for a particular child and the sheer number of records to review could take the entire length of the visit to count. This was particularly true in classrooms that had frequent child data collection. The first three rounds of visits occurred in late spring 2015 (rounds 1–3), 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,

³¹ In the first two visits, we asked for the evidence that they collected for that child without specifying a time period. Teachers often provided us with evidence that was too extensive to be able to evaluate (for example, spanning the entire program year or several program years). We also considered the fact that the EDIT might be used in fall and spring so we limited the time period to the two months prior to our visit.

and even counting the number of anecdotal records within each developmental domain could consume much of our time in the setting and the number of anecdotes did not reflect the quality.³² That is, teachers who wrote down more anecdotes did not necessarily have higher quality evidence. Many observations were not focused and the accompanying documentation did not provide clear evidence of children's skill levels.

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 subsequent visits, we sampled the previous two months of documentation and reviewed some data related to the child's initial 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 the 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 rarely received the exact sequence requested. Our revised written instructions did sometimes produce the exact number of videos requested; however, the content still differed slightly from what we had asked for. A call with the teacher to walk through our video request could be helpful, but most teachers did not take us up on our offer of a call.

C. What are the implications of the pretest for the recommendations of future procedures?

In general, teachers reported that participating in the EDIT was not overly burdensome, taking them approximately three hours spread across several weeks to collect the necessary documentation, video record themselves, and participate in the teacher interview. Throughout the pretest, we continued to refine our communications with teachers to ensure that the materials

³² Even when we asked teachers to limit the data they shared with us to a specified time period, some gave us all the data they had for the focal children.

needed could be assembled without giving the impression that participation would be overwhelming. The original letter was long and detailed (three pages of information and a two-page tracking sheet). During initial 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 (Appendix C). We also revised the language to avoid confusion about what we meant 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 two focal children.” We also clarified that we were not observing the classroom *in person* during our visit, so the focal children did not have to be present in school on the day of our visit.

Based on our pretest experiences, we recommend five procedures for future implementation of the EDIT: (1) gaining consent for all children while asking the teacher to select the two focal children; (2) using succinct 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 to video record and the data needed for 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 raters can meet with the teacher in the classroom to provide the teacher with 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 holistic rubrics, analytic rubrics, ratings, and 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.

In another example, the interview helped clarify that, although one teacher made minimal use of the commercial ongoing assessment system other than as a repository of information grouped by child, the teacher was collecting data outside of the system. That teacher was also engaging in daily meetings for data interpretation with co-teachers who took notes on observations of children each day. Therefore, although the use of the formal system and documentation from that system was minimal, the interview revealed that the teacher was thoroughly engaged in a locally created system of assessment, interpretation, instructional planning, individualization, and reassessment. The video examples also demonstrated the procedures that the teacher described in the interview.

Overall, we recommend retaining all three data sources. This allows raters 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 G for additional data sources we considered).

E. How did the EDIT items and interview protocol change during the pretest?

Two researchers from the EDIT study team rated each classroom. Researchers teamed together in different pairs across the observations. Each researcher rated independently; the pair then 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 study team met to discuss the disagreements and areas that were challenging to code.

Changes made to the EDIT instrument during the pretest included:

- Refining the wording on the rubrics (for example, clarifying that the term “progress” meant “at least three times per reporting period”)
- Refining the criteria within rubrics and ratings (for example, making sure that criteria were described in similar ways across levels within a rating scale to promote consistency)
- Changing some items to binary checklists (yes or no) or a four-point rating rather than a seven-point rating
- Changing some ratings to the frequency of use, rather than tying them to the number of assessment targets
- Adding skip logic to the instrument
- Adding options for “Not Applicable” and “Cannot Rate” for select items
- Removing items that were redundant (we tested different ways of assessing constructs and selected the one that was most efficient and captured teacher practice)
- Changing the flow of the questions in the teacher interview and adding a cover page of “highlights” to assist the interviewer
- Adding questions and prompts to the teacher interview
- Collecting definitions and examples for key ideas in the items to share among the EDIT raters (Appendix E)

1. Items with limited variation

EDIT items with limited or no variation in scores 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, Inc. 2013). After the full team debriefed, 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 one to four scale (rather than one to seven) and asked raters to note whether the item was based on information gathered from all data sources.

During the third round, two EDIT raters visited a program that had not provided training to the teacher on 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 variation 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 variation in our earlier pretest visits, we tried both the one-to-four characteristic rating scale items and binary items when evaluating the documents, videos, and interview responses.

We found that items using a one-to-four 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 ninth 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 ninth visit, we changed some of the binary items to the four-point rating scale. We tested these items in nine additional classrooms and found variation among the teacher's scores.

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

Item	Hypotheses for lack of variation	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.	Examined in a more diverse set of classrooms; probed more about what the teacher was doing when anecdotal records and running records were being collected; however, variation remained low (likely because we pretested in classrooms using curriculum-embedded ongoing assessments).
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.	Changed to a 1 to 4 rating, noting if the item is strongly characteristic, sought consent from all children in the classroom to facilitate video recording in classroom environment. Variation improved after these changes.
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; however, variation remained low (likely because we pretested in classrooms using curriculum-embedded ongoing assessments).
3d. 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. Variation improved after this change.
3f. 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. Variation improved after this change.
3g. 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 intentionality rather than efficiency. Variation improved after this change.
4g. Organization facilitates communication with families	TS GOLD and WSS both offer a variety of types of graphs that the 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. Variation improved after this change.
4p. Organization facilitates comparisons	TS GOLD and WSS both offer a variety of types of graphs that the 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. Variation improved after this change.

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

Based on our first nine pretest visits, among the study 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 five 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 one, two, and three grapes right away, usually counted correctly when there were four grapes, and when there were only five, counted the same grape twice, reporting six 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 five, she showed a group of three with a group of two; or a group of four with one; or a line of five. She planned to give this child four or five 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 skill was improving, she might continue using that strategy for the next week with different objects (such as counting bears, blocks, or 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 during 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 final four pretest rounds (rounds 3–6), 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 and rating scale items, and we added to the criteria for several of the rubrics. We found that the rating scale items with a four-point scale were the easiest to rate. The additional criteria made it more difficult to attain the highest rating on the revised rubrics. Subsequently, the two raters had exact agreement on the independent ratings of almost 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	<p>This criterion was added to receive a score of 5: <i>At least one assessment target is represented in individualized plans for instruction for at least some of the children.</i></p> <p>This criterion was added to receive a score of 7: <i>At least one assessment target is represented in individualized plans for instruction for all children.</i></p>
Increased measurement of progress and the effects of individualization	
1	This criterion was added to receive a score of 5: <i>At least one target per child is measured over time to track progress.</i>
2b	A rating item was added that states: <i>Only 1 or 2 assessment targets are assessed more than once per child to examine progress.</i>
2k	A rating item was added that states: <i>Teacher monitors child's progress in area of individualization with at least 3 pieces of evidence (data points).</i>
2l	A rating item was added that states: <i>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.</i>
2m	A rating item was added that states: <i>Teacher continues to periodically monitor child's progress in area of individualization even after first signs of improved progress.</i>
3h	A rating item was added that states: <i>Documentation includes child's responses/performance during individual learning plan activities.</i>
4a	A binary item was added that states: <i>The teacher views/organizes the data to compare a child's performance to a developmental expectation or benchmark for growth.</i>
4b	A binary item was added that states: <i>The teacher views/organizes the data to compare a child's progress to a developmental expectation or benchmark for growth.</i>
4c	A binary item was added that states: <i>The teacher views/organizes the data to compare a child's performance to that of other children in the class.</i>
4d	A binary item was added that states: <i>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.</i>
4e	A binary item was added that states: <i>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).</i>
4h	A binary item was added that states: <i>The teacher organizes the information to look at performance by class for one or more assessment targets at a single time point.</i>
4i	A binary item was added that states: <i>The teacher organizes the information to look at progress by class for one or more assessment targets across multiple time points.</i>
4j	A binary item was added that states: <i>The teacher organizes the information to look at performance by subgroup for one or more assessment targets at a single time point.</i>
4k	A binary item was added that states: <i>The teacher organizes the information to look at progress by subgroup for one or more assessment targets across multiple time points.</i>
4l	A binary item was added that states: <i>The teacher look at performance by domain for the class or a subgroup for one or more assessment targets at a single time point.</i>

Item	Example of revision
Increased measurement of progress and the effects of individualization (cont'd.)	
4m	A binary item was added that states: <i>The teacher organizes the information to look at progress by domain within reporting periods (multiple assessment targets across multiple time points).</i>
4n	A binary item was added that states: <i>Teacher uses visual displays to depict child progress by date for at least one target.</i>
4o	A binary item was added that states: <i>Teacher indicates in documentation when a new instructional strategy or individual learning plan is implemented.</i>
5a	This criterion was added to receive a score of 5: <i>Teacher examines more than one data point in between checkpoints for at least one target per child (such as preliminary ratings).</i>
5b	Additional guidance was provided. As an example, the criteria to receive a score of "7" specified: <i>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.</i>
6a	This criterion was added to receive a score of 7: <i>Considers child's progress relative to typical development and progress of peers.</i>
6i	A rating was added that measures whether evidence is used make decisions about instructional strategies.
7	Following criteria were added to receive a score of 7: Plans and collects at least 3 pieces of evidence about how each child responds to individualized goals and instruction. ; <i>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).</i>

Note: All changes are made in the final pretest version of the EDIT.

We also added five 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.

Table V.5. Additional EDIT items

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

Note: All changes are made in the final pretest version of the EDIT.

Based on a recommendation from ACF to obtain more information about the assessment context, we asked teachers during the last 10 visits 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 for this information.

3. Use of “don’t know” ratings

During the first three rounds of the pretest (pretest rounds 1–3), 19 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. We recommend that future users of the EDIT make sure that the requests for information from teachers use terms that are familiar to teachers using that particular assessment system. In our final nine pretest visits, EDIT raters did not rate any items as “don’t know.”

4. EDIT instrument conventions

During the final three rounds of the pretest (pretest rounds 4–6), we implemented several changes to the EDIT to make it easier for raters to complete the EDIT ratings without having to consider items that did not pertain to a particular classroom. For example, we grouped several questions about a teacher’s use of a data collection plan together and preceded that section with the gate-keeping item: *Does the teacher have a plan for data collection?(Yes/No)*. If the answer was “Yes,” the rater would answer the related questions, and if the answer was “No,” the rater would skip those items and proceed to the next section. In addition, because some items apply only in specific circumstances, we added rating options for “Not Applicable” and “Cannot Rate” to several items. For example, item 2H applies only when a teacher uses a standard task assessment (if there is no standard task, it is rated as “Not Applicable”), and it can only be rated when documentation for that standard task is available (if no documentation is available, “Cannot Rate” should be selected).

5. Interview protocol

During the pretest, EDIT raters conducted a semistructured interview with the lead classroom teacher. The interviewers probed for explanations about documents and observations gathered, as well as to obtain information on the teacher’s planning and implementation of instructional adaptations and individualized teaching strategies. Because the interview questions were open-ended, respondents sometimes gave responses that were pertinent to several interview questions and topics at one time. Because the conversation often deviated from the order in the interview protocol, this made it difficult for the raters to track whether all important topics were covered. To support the EDIT raters, we added a two-page cover sheet to the interview protocol (Appendix A) to highlight the key topics the interviewer should touch on to ensure all essential information is covered during the interview while maintaining a semistructured approach that eliminates redundancy. Interviewers referred to the cover sheet throughout the interview for the final six pretest visits; they noted that it was easy to use and helped ensure that the full set of

relevant topics were covered during the interview. Throughout the pretest, we also sometimes revised or added to the questions in the interview protocol. For example, we added a question to the teacher interview to help elicit information about reviewing data by subgroups to be able to answer EDIT instrument item 4j (*The teacher organizes the information to look at performance by subgroup for one or more assessment targets at a single timepoint*).

6. Training materials

Before our final nine visits, we trained two additional raters from outside the original EDIT study team. Before the training, trainees reviewed the EDIT instrument, interview protocol, and item-by-item explanations and examples (Chapter II, Appendices B and F). The training was conducted in three three-hour blocks and included a description of the project, a description of the EDIT data sources, EDIT item-by-item explanations (including more than 30 video and photograph examples gathered from previous visits), a description of the EDIT interview, and general site visit procedures (Appendix E). Before the final training session, trainees were asked to score the EDIT using pictures of documents, videos with focal children, and a recording of a teacher interview from a single classroom. During the final training session, scores from that exercise were discussed. The two additional raters accompanied experienced EDIT study team members into the field, and scoring agreement among raters remained high. For additional information about training, see Appendix D.

7. 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 ongoing assessment 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 and tested these items during the final three rounds of the pretest (rounds 4–6), 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. We tested these revised items in the final nine classrooms and found variation in teachers' intentionality and use of data to inform decisions. The final EDIT measure is better able to capture a range of competencies in teachers' use of ongoing assessment to inform instruction.

The pretest was staggered in six rounds 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. In the next chapter, we propose next steps for the EDIT.

Key findings

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

- All reported they would recommend participating in the study to colleagues.
- Some teachers reported that the project looked overwhelming when first receiving our materials, but they reported minimal burden after they got started (3 hours).
- 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; will allow teachers to record small group instruction.
- Using revised written materials to facilitate communication and to not overwhelm the teacher; will help to build cooperation between researchers and participants.
- 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; will ensure materials are prepared according to specifications.
- Requesting a copy of the program's edition of the ongoing assessment system for use during the visit; will assist researchers in understanding the data that is collected.
- Whenever possible, scheduling the teacher interview for a time when we can meet with the teacher in the classroom; will provide the teacher with easier access to a full range of materials and assist with teacher recall.

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 pretest 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 variation.
- Adding conventions to the instrument to support raters' completion of the instrument (for example adding "Not Applicable" as an option for some items)

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

- Creating a cover sheet to highlight key concepts to support the interviewer in administering the semistructured interview
- 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

In this chapter, the study team proposes some next steps for obtaining evidence of the psychometric properties of the EDIT to support use of the EDIT as a research instrument. We also discuss the challenges to this work. The chapter concludes with a discussion of the study team's vision of the potential future uses of the EDIT.

A. Recommended next steps for further work

Only 18 classrooms were assessed with the EDIT during the development phase. More testing is needed to assure a valid and reliable measure. To further examine the psychometric properties of the revised EDIT, the study team recommends a pilot test. The goals of a pilot test could include the following:

- Testing the EDIT with a broader set of classrooms, looking for sensitivity to differences in practices and teacher backgrounds
- Evaluating the reliability of the measure, including both internal consistency and interrater reliability
- Evaluating different approaches to scoring the EDIT (for example, providing scores by dimensions and then averaging for a total score; weighting different items)
- Examining evidence of validity, including convergent and discriminant validity
- Continuing to expand operational definitions, as needed, to apply to a broader sample of practice

Sensitivity to teacher and context differences in practice. We recommend testing the EDIT with a broader set of more diverse classrooms and looking for sensitivity to differences in teacher practices—that is, does the measure detect differences in how teachers implement ongoing assessment to inform instruction and individualization? This would include looking at the variation we find in the items and whether the difficulty of implementing assessment practices is consistent with our theoretical expectations that the most challenging items involve interpreting data (both current performance and progress) and selecting, implementing, and monitoring the success of evidence-based strategies that are aligned with the needs of individual children. We propose sampling a broad set of classrooms with different (1) assessment systems, (2) levels of teacher education, (3) access to technology, and (4) levels of coaching and administrative support. We hypothesize that this variance in context will affect how teachers implement assessment practices. For example, we expect that teachers who have administrative support and easy access to online systems and evidence-based or professionally recommended instructional practices will use data more frequently than teachers in classrooms with fewer supports.

Evaluating reliability and scoring the EDIT. Analysis should include examining the reliability of the items (internal consistency) and the dimensionality (the different constructs/factors measured) in the EDIT. Researchers could consider whether there is a need to combine items for scoring purposes (perhaps by assessment activity) or weight some items more strongly in assigning scores. For example, scoring might give more weight to the items that

reflect skill in interpreting data or selecting and monitoring the success of evidence-based instructional strategies aligned with child performance. These items are critical for reaping the benefits of assessment for instructions. We anticipate that researchers would calculate scores by dimension (for example, scoring the different stages in the conceptual model such as the quality of the teacher's focused observations, objective and complete documentation of key evidence, unbiased interpretation of evidence, selection and application of appropriate instructional strategies).³³

A pilot test would allow examination of the inter-rater reliability of the EDIT when independently scored by two raters. Because some assessment systems are more transparent and provide more information for a rater to use in scoring the EDIT, the reliability of the raters may vary by assessment system and familiarity with the system. We recommend examining inter-rater reliability overall and also looking for any patterns in rater agreement by assessment system.

Examining validity. A measure that is valid is measuring what it is supposed to measure. Initial content and face validity evidence is already available. The EDIT draws on the literature review we conducted at the start of the project. We also had our expert panel review and confirm the face and content validity of the items. With a larger sample size, we recommend the use of confirmatory factor analyses and/or item response theory (IRT) analyses to look at construct validity.

Analyses of convergent and discriminant validity provide evidence of what the EDIT is (convergent) and is not (discriminant) measuring. Obtaining evidence of convergent validity will be the most challenging. To evaluate convergent validity, researchers typically would look for associations with other measures of the same construct and/or outcomes that are expected to vary based on the teacher's implementation and use of ongoing assessment. However, identifying measures or related constructs to use in obtaining evidence of convergent validity for the EDIT will be difficult, because our literature review indicated that a measure similar to the EDIT does not exist.

Weak, negative, or absent relationships with other constructs provide evidence of discriminant validity. For example, although strongly skilled teachers may do well across all aspects of teaching, we would expect a weaker relationship between health and safety practices and scores on the EDIT than between instructional quality and the EDIT. Based on our observations, we recommend that the different dimensions of the EDIT (quality of data collection compared to quality of interpreting data) be examined in relation to outcomes separately and posit that there may be interactions with teacher experience and education. Although teachers with more experience are highly observant of child behavior, they may more narrowly target the information they record and thus collect less data than less experienced teachers. On the other hand, teachers with less experience and expertise may need to record more data to have the data benefit their instruction.

³³ Because selecting the assessment target and conducting the assessment are interrelated, and our measurement of these is intertwined (for example, the "analytic rubrics and checklists for selecting and implementing the assessment method"), these would continue to be scored together.

Definitions and examples to create additional training materials. Clear definitions of terms and processes would support the use of the EDIT by a variety of raters. Examples also help ensure that different raters interpret the rating rubrics in the same way. During the 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,” and tracking progress was described as “at least three pieces of evidence within a reporting period.” 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. (These definitions and examples are represented in Appendix E.) Additional studies with a much larger sample could continue to expand these materials with experiences in classrooms using different assessment systems or those taught by teachers from differing backgrounds or serving different populations.

1. Challenges to future work

As noted, identifying a measure or indicator that could be used to examine convergent validity of the EDIT will be a key challenge. The review of the literature did not reveal any other measures of teacher implementation of assessment for individualization that we could use for this purpose. A related construct that we would hope to observe is an increase in children’s rate of progress in classrooms with strong implementation of ongoing assessment. Because the teachers administer all the classroom assessments in most centers, the pilot test would need to independently collect fall and spring (summative) assessment data to avoid this confound. The measure of child outcomes would need to be sensitive to change in a brief period of time, and a child outcomes data collection would increase both burden and the expense of a study.

Other constructs that could be used in evaluating the validity of the EDIT are not as closely aligned and would likely have weak to moderate relationships. For example, teacher education or experience in conducting and using assessments, or observational measures of the quality of teacher interactions, could provide indicators of teacher knowledge or teaching skills, but they will not capture actual assessment practices. Frequency of and fidelity to professional development and mentoring in assessment will hopefully provide moderate associations with the EDIT. Strong assessment practices should help teachers become more familiar with the children they teach and increase skill in scaffolding child success. This could help teachers and children to have more positive relationships and provide an additional construct to examine in relation to the EDIT. However, there are other factors that will influence teacher-child relationships and so correlations may be weak. Detecting weak to moderate correlations requires a large sample.

2. Sample size

In designing a pilot test of the psychometric properties of the EDIT, we need to consider the goals of the pilot test, the expected strength of associations, and potential sources of error that could weaken the ability to detect the associations. Weighing these factors helps determine what we might gain or risk with different sample sizes. A small sample, such as 40 additional classrooms would be adequate to gather some psychometric properties but could limit what a researcher would be able to detect. In order to detect evidence of reliability and validity, the measure would need adequate to strong reliability across the different assessment systems used in the sample and moderate to strong relationships with related constructs. For measures of internal consistency, such as alphas, the reliability estimates may not be as precise. One study

indicated that high reliability with 30 to 40 people (with about 35 items) decreased as sample size increased until the sample reached about 400 (Charter 1999). If the reliability is good, the estimates will stay within an acceptable range. Charter recommended more than 400 for stable estimates, but found that 59 percent of published studies estimated reliability based on samples of fewer than 100. When you have fewer items (for example, subscales with 4 to 5 items), items need to have a stronger item-to-total correlation to achieve high reliability. A small pilot test could provide initial evidence of internal consistency and explore evidence of validity. However, if the measure has low reliability, a small sample size will limit the ability to detect significant relations among measures and thus limit validity evidence.

With adequate reliability, a sample size of 40 classrooms (internal consistency $>.80$) would allow detection of moderately strong correlations (0.75 power to detect r of 0.40 and 0.86 power to detect $r = > 0.45$) with other constructs that we expect will relate to the use of ongoing assessments to monitor progress and individualize instruction. As noted above, however, we lack strongly aligned measures and expect to find only weak to moderate correlations.

Because the available assessment systems offer varying levels of support³⁴ for teacher implementation and interpretation, the assumption of good to strong internal consistency across assessment systems may be overly optimistic. One assessment system might provide explicit guidance about what data to collect and provide forms on how to collect the information (supporting higher ratings on the teacher implementation of these stages of assessment). Another assessment system might provide little guidance on what or how to collect data but might organize and aggregate data within domains to aid the teacher's interpretation. The items associated with these stages of assessment may contribute to the measurement of teachers' use of assessment data in very different ways across these two assessment systems. This would lead to lower internal consistency. Although these areas could be examined as subscales, it is more difficult to obtain reliability with brief scales. The individual items within brief scales need high item-to-total correlations for a reliable scale.

A sample size of 40 also would not allow examination of subgroups. Strengths and weaknesses of practice may differ across teachers with different levels of training and education. Differences in knowledge and implementation of practices could make it difficult to obtain reliable scores when looking across subgroups. As noted earlier, the reliability of an instrument affects the resulting power to detect relationships (Raudenbush et al. 2008) and thus would also affect the ability to find evidence of convergent validity.

Therefore, to examine the psychometric properties of the EDIT, the study team recommends sampling at least 100 additional classrooms, as follows:

- At least 30 classrooms for each of three types of assessment to examine internal consistency, rater agreement, evidence of construct validity, and dimensionality

The study team also recommends that the pilot test continue to limit the sample to classrooms using curriculum-embedded assessments that examine multiple child domains and have at least one peer-reviewed article with evidence of psychometric properties. Assessment

³⁴ In addition, programs may offer varying level of training and support for teacher's implementation of the systems.

systems used by Head Start classrooms, as reported in the PIR, may provide a strong starting point for selection. In the pretest, we noted some similarities across curriculum-embedded assessment systems. Teachers are asked to collect multiple pieces of evidence about children’s development. Behavioral descriptions are provided for the criteria to be met for each dimension within the domains assessed. Most assessment systems use anecdotal records and samples of children’s work as evidence of their skill, knowledge, or behavior. Most systems have an online component available with summary reports or graphs (although these features sometimes are optional).

The pilot test could expand prior work by including classrooms that have different levels of teacher education and that have not purchased the online data entry and supports. Purposive or stratified random sampling of classrooms would ensure diversity of classrooms and at least 30 classrooms that did and did not have characteristics that might affect assessment practices (Table VI.1). For example, teachers with larger class sizes would likely have more difficulty collecting data on all the children and all the domains. Researchers might sample at least 30 large and 30 small classrooms³⁵ to allow examination of how class size affects ratings on the EDIT.

Table VI.1. Sampling selection criteria options^a

	Minimum number of classrooms targeted	
	With	Without ^b
Technology support	30	30
Aligned with Classroom Curriculum		30
Supplementary materials for intentional, targeted lessons	30	
B.A. or greater education level	30	30
Head Start (versus state-funded pre-K)	30	30
Class size greater than 14 ^c	30	30
Instructional or assessment coaches or mentors	30	
Mixed-age group (versus only 4-year-olds or only 3-year-olds)	30	30 (single age; either age is fine)
Community-based program (Head Start or private versus programs in schools)	30	

^aThe full sample size would be 100 classrooms with at least 30 in each block of classrooms.

^bBlank cells would not have a sampling requirement.

^cMean classroom size according to FACES classrooms observations in spring 2010 was 14.2 children.

We recommend including at least 30 classrooms using the assessment most frequently used in early childhood settings, (currently, this is the TS GOLD with and without technology),³⁶ 30 classrooms using another assessment *with* a technology component that organizes or interprets the data for the teacher, and 30 classrooms using an assessment *without* the technology component. In our review of the literature, the studies that used assessments with a technology component were more likely to have evidence of effectiveness (Buzhardt et al. 2010, 2011; Landry et al. 2009). Additional research that focuses on key elements of technology components

³⁵ Large and small class sizes will be based on the average class size for Head Start programs. Large classes will be those with more than 14 children enrolled in the class.

³⁶ At least half of the classrooms in Head Start that use TS Gold use the online version.

and how they support teachers in using data for instruction would add to the field's knowledge of what is needed for effectiveness.

The other assessments could be selected from the assessments used in early childhood settings that have evidence of validity (Table VI.2) and that have some features that overlap and some features that contrast with one another. For example, in our pretest, we selected WSS because it has a technology component (similar to TS GOLD) but is not directly aligned with any curriculum and is preschool to grade 3 rather than birth through preschool. Researchers might select another measure that is also birth through preschool and aligned with a curriculum, such as the Child Observation Record-2 (COR-2) or the Assessment Evaluation and Programming System–Second Edition (AEPS-2) but that does not use the technology component of the COR Advantage or AEPS*interactive*. To answer questions about technology components or different features of an assessment system, the teachers selected for the study should have at least one year of experience with the measure. The study team recommends enough variety of assessments to capture the range of practices, but researchers also must keep in mind that raters will have to become somewhat familiar with the assessments' components and terminology to ask for needed information and understand how the teacher talks about the data (teachers often adopt the abbreviations and terms the assessment system uses).³⁷ Having features overlap with some assessments but not others could help to examine how that feature (for example, provision of checklists, recommendations for activities, multiple examples of how a child might exhibit a skill, guidance for portfolio collection) supports the assessment process and whether the EDIT is sensitive to the differences in teacher practices across assessments systems with and without that feature.

³⁷ During training and before each visit, raters should review the assessment manuals for the system that they will encounter on the visit.

Table VI.2. Commonly used ongoing assessments as reported in the 2014–2015 Program Information Report

Assessment	Age/grade range	Percentage reporting in PIR	Validity evidence in peer-reviewed publications
AEPS; AEPS-2nd edition	Birth to 5	<1	Bricker et al. 1990; Gao and Grisham-Brown 2011—evidence of concurrent validity and sensitivity; Macy et al. 2005; Hallam et al. 2014
Child Observation Record (COR, COR-2, or COR Advantage)	6 weeks to 6 years	7.0	Fantuzzo et al. 2011 (COR) and Barghaus and Fantuzzo 2014 (COR-2) found problems with the COR and COR-2 factor structure and scale; Fantuzzo et al. 2002 and Fantuzzo et al. 2011 – some support for convergent validity; Sekino and Fantuzzo 2005 found convergent and predictive validity for COR
DRDP-R; DRDP Access	Birth to 5	5.2	Karelitz et al. 2010
E-LAP, LAP, and LAP-D*	Birth to 6	5.1PIR	NA ¹ ; MacMann and Barnett 1984, and Barnett et al. 1988 validity evidence not met
Galileo	Birth to 5	3.5	NA ¹ ; developer reports inter-rater reliability and internal consistency
<i>Teaching Strategies</i> GOLD* or Creative Curriculum Developmental Checklist	Birth to 6	59.5	Lambert et al. 2014; Kim et al. 2013; Kim and Smith 2010
Work Sampling System	Preschool to grade 3	2.0	Meisels et al. 2001; Meisels 2003

¹NA “Not Available”. We did not find evidence of validity in peer-reviewed publications.

Across the classrooms, a range of teacher education and background (training and experience) in conducting assessments would allow researchers to examine associations with education and experience. Analyses of associations may require controlling for or analyzing assessment systems separately. Some assessment systems are more supportive than others and may have low correlation with education and experience, and other assessment systems require greater teacher knowledge and experience. All the teachers should have received some training in the assessment they are using.

3. Recommended procedures for pilot test

The training materials created in this project (Appendices D, E, and F) could be used as a starting point for developing EDIT rater training. In addition, researchers would need to provide some training for raters about the key features of the selected assessment systems and the terminology they may need to know to be able to prompt teachers (during the interview) about their practices using a particular assessment. The rater needs to be familiar with the terms used in the assessment to communicate more easily with the teacher about what is needed for EDIT document review and to tailor questions in the interview.

Raters should demonstrate reliability in using the EDIT before collecting data independently. After reliability certification, we recommend sending two raters per classroom for at least 10 percent of classrooms for each assessment system, and examine inter-rater reliability based on these independent ratings. After completing the visits, raters could discuss any

discrepancies to determine whether additional clarifications are needed to rate practices used with a particular assessment system. These visits should occur throughout the field period to check for rater drift.

The study team recommends conducting debriefing calls with raters weekly or biweekly, depending on the frequency of classrooms visits. These calls could help clarify anything that raters encounter in the field that was not anticipated and provide additional information that will help refine wording and guidance in future training materials. These calls could also ensure that raters are using the EDIT in consistent ways and documenting the information necessary to justify their scores.

4. Validity evidence

As noted earlier, initial evidence of content validity is available from our research base in the development of the EDIT and from the review and feedback from our expert panel.

To obtain evidence of construct validity, we recommend using a Rasch rating scale or partial credit model to analyze the data. Stable estimates can be obtained with smaller sample sizes (for example, 100) than what is needed for confirmatory factor analyses. Given the different types of items (for example, ratings from seven point rubrics and dichotomous ratings), researchers would need to use the “groups” command in analysis to account for the differences in the range of possible scores. Researchers would need to examine the dimensionality to determine whether a single or multiple scales are needed to account for the variance in the data. We recommend examining the residuals with principal components analyses to determine whether there are multiple dimensions that should be measured separately.

To obtain some evidence of convergent validity, one possible approach is to examine the teacher’s EDIT rating in relation to the amount of professional development (including coaching and mentoring) in conducting and interpreting assessments, as well as associations with other selected indicators of interest. (See Table VI.3 for advantages and disadvantages of different indicators for examining validity.)

Table VI.3. Advantages and disadvantages of constructs considered for validating the EDIT

Constructs of interest	Rationale	Advantages	Disadvantages
Child performance and progress			
Child progress—based on independently administered assessments	Goal for conducting ongoing assessment and individualization	Most closely aligned with project goals	Increased burden Expensive to collect Need large sample sizes to look at gain scores (more error in gain scores than in a single time point)
Variability in teacher ratings	Teachers who assess regularly will be more sensitive to the strengths and weaknesses of different children.	Could potentially collect from administrative records (particularly when programs use a computerized system) Provides a link to child outcomes	Dependent on the range of variability in the assessment that the teachers use Adds burden and expense of obtaining consents from all parents
Teacher self-reported information			
Teacher education in early childhood/child development	Teachers with more education should be more knowledgeable about child development, assessment, and evidence-based practices.	Inexpensive to collect	Degree is more distal (to knowledge of preschool assessment and instruction) and does not tell how many courses in particular areas. Teachers often do not remember how many courses they had in child development or assessment. Even courses with similar names—"assessment" or "development"—may have different emphases on preschool assessment and development.
Teacher knowledge of assessment and child development (test)		Brief test would be relatively inexpensive to collect (could do as web survey).	Additional teacher burden Reliable and valid measures of assessment knowledge are not currently available.
Administrative or parent ratings of teacher			
Supervisor ratings of proficiency in assessment	Supervisors should be familiar with both the assessment system and with each teacher's typical practice.	Inexpensive to collect; relatively low burden Ratings may be more representative of typical practice (rather than the video-recorded observations the teacher provides), but are likely based on different levels of familiarity with the teacher's use of data.	Supervisors may be biased in their ratings. The SAQ and rating scale would need to be developed and would not have preexisting evidence of reliability and validity.

Constructs of interest	Rationale	Advantages	Disadvantages
Parent ratings	Parents would know what types of information the teacher shares with them and whether the teacher seems to know their children's strengths and weaknesses.	Low burden Can obtain the perspective of multiple parents in the class, increasing the reliability of the ratings; biased—reliability differs across socio-economic levels.	More expensive to collect, with a need for permissions and responses from multiple parents; timing will matter—parents who just had a quarterly review or end-of-semester meeting will have more knowledge about teacher practices than other parents Some teachers may not involve parents in assessment at all.
Classroom quality ratings			
CLASS dimension scores	Theoretically, teachers who use assessment data should have stronger instructional support for children.	Data are already being collected in many classrooms.	Limited variance will negatively affect power to find relation. The range of scores on Instructional Support on the CLASS is very limited nationally. In addition, some teachers will have positive instructional support but will not individualize and may have weak assessment skills. We would not expect a strong correlation.
ECERS-R subscales scores	Certain environmental scales (e.g., Personal Care routines) should be weakly related to the process evaluated by the EDIT. This could be used as evidence of discriminant validity. Teacher interaction items would help in testing convergent validity. Hypotheses about the types of environments that support assessment to inform instruction could also be tested.	Data are already being collected in many classrooms.	Additional cost to collect this information. The interaction scales are not well aligned with assessment and individualization. Some teachers will have positive interaction with children, but will not individualize and may have weak assessment skills. We would not expect a strong correlation.

B. Potential future uses

After additional psychometric examination and refinement, the EDIT could be used in the future for research and possibly also inform the design of a professional development program or help in examining necessary program supports. The EDIT has been presented at research conferences—the literature review and measurement plan in particular—and these presentations have led to interest in the field both to learn more about the measure and to potentially use it. The EDIT has the potential to 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,³⁸ 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 what practices are being used in the field, how much variation there is in teacher practices, and which practices are most strongly associated with positive outcomes for children.

2. Professional development

The collection of evidence from the pretest 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. From the guidance provided in the rubrics, the strengths and weaknesses in the teachers' assessment practices were clear in the 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 might be used to develop professional development programs to support teachers and then evaluate teachers' use of assessment for individualization in programs.

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

Any evaluative use of a classroom measure 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. The pilot test proposed in this chapter could provide information on features of assessments that are needed to support teachers in planning, collecting, and using assessment data. Collecting additional information about context (for example, administrative support, coach availability, professional development time, and culture of assessment) could help answer additional questions, such as: How much time do teachers need for professional development that focuses on implementation of the assessment system in use in their center? What information do they need to be successful in using assessment data to inform instruction? 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. Depending on the questions of interest and theory about ongoing assessment (or evidence from

³⁸ For example, technology supports that provide immediate recommendations to teachers have evidence of a positive association with child outcomes.

psychometric work that only some dimensions are critical), researchers may choose to focus on only some of the scales within the EDIT.

Key considerations for the pilot test described in this chapter

The primary goal of the pilot test would be to obtain psychometric evidence of the EDIT in research:

- Sample needs to be selected purposively or use a stratified random sample to ensure at least 30 sample members in each key subgroup.
- Sample size needs to assure the power to detect weak to moderate correlations.
- No single instrument is identified as ideal for assessing convergent validity of the EDIT. Selection of an appropriate indicator for gathering evidence of convergent validity will require weighing the advantages and disadvantages of different criteria.

The pilot test could provide:

- Additional information on how characteristics (for example, class size, classroom composition) of the classroom or program are related to differences in how teachers implement assessment for instruction and individualization
- Some information on how teacher assessment and individualization practices vary by characteristics of assessment systems
- Training examples of how teachers using other assessment systems implement the practices in the EDIT

The primary recommended use of the EDIT at this time is:

- Psychometric research to collect evidence of validity
- Descriptive research to learn about what teachers are doing to collect and use assessment data for instruction and individualization

After additional research has been conducted to more carefully examine the measure's psychometric properties, the EDIT could be used for the following:

- To identify how different practices within the EDIT contribute to assessment and individualization, and use that information to develop professional development programs to support teachers
- To examine and inform how programs support teachers in conducting and using assessment data

REFERENCES

- Administration for Children and Families. "Head Start Performance Standards." 45 CFS Sec 1302.102.3(c)(2)(ii). *Federal Register*, September 1, 2016.
- Administration for Children and Families. "Head Start Early Learning Outcomes Framework." Washington, DC: ACF, 2015. Available at <https://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach/pdf/ohs-framework.pdf>. Accessed July 26, 2016.
- 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, S., C.M. Connor, J.S. Folsom, L. Greulich, J. Meadows, and Z. 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, S., E. Moiduddin, L. Tarullo, M. Caspe, S. Prenovitz, and A. Jewkes. "The Learning from Assessment Toolkit." Washington, DC: Mathematica Policy Research, March 2009.
- Bagnato, S.J., M. McLean, M. Macy, and J.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, S.J., J.T. Neisworth, and K. 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.

- Barghaus, K.M., and J.W. Fantuzzo. "Validation of the Preschool Child Observation Record: Does It Pass the Test for Use in Head Start?" *Early Education and Development*, vol. 25, no. 8, 2014, pp. 1118–1141.
- Barnett, D.W., J. Faust, and M. Sarmir. "A Validity Study of Two Preschool Screening Instruments: The LAP-D and DIAL-R." *Contemporary Educational Psychology*, vol. 13, no. 1, 1988, p.p. 26–32.
- Bodrova, E., and D.J. Leong. *Tools of the Mind: The Vygotskian Approach to Early Childhood Education (2nd ed)*. Upper Saddle River, NJ: Prentice-Hall, 2007.
- Bolt, D.M., J. Ysseldyke, and M.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.
- Bricker, D.D., E.J. Bailey, and K. Slentz. "Reliability, Validity, and Utility of the Evaluation and Programming System: For Infants and Young Children (EPS-I)." *Journal of Early Intervention*, vol. 14, no. 2, 1990, pp. 147–160.
- 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.
- Buysse, V., E.S. Peisner-Feinberg, E. Soukakou, A. Fettig, J. Schaaf, and M. Burchinal. "Using Recognition and Response to improve children's language and literacy skills: Findings from two studies." *Early Childhood Research Quarterly*, vol. 36, 2016, pp 11-20.
- 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.

- Carta, J.J., C.R. Greenwood, J. Atwater, S.R. McConnell, H. Goldstein, and R.A. Kaminski. "Identifying Preschool Children Who Need More Than Tier 1 Language and Early Literacy Instruction." *Journal of Early Intervention*, vol. 36, no. 4, 2015.
- Carter, D.R., and R.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.
- Charter, R.A. "Sample Size Requirements for Precise Estimates of Reliability, Generalizability, and Validity Coefficients." *Journal of Clinical and Experimental Neuropsychology*, vol. 21, no. 4, 1999, pp. 559–566.
- 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.
- Classen, A. and G. A. Cheatham. (2015). "Systematic Monitoring of Young Children's Social-Emotional Competence and Challenging Behaviors." *Young Exceptional Children*, vol. 18, no. 2, 2015, pp. 29-47.
- 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, B.D., D.M. Gorecki, and L. 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.
- Fantuzzo, J., D. Hightower, S. Grim, and G. Montes. "Generalization of the Child Observation Record: A Validity Study for Diverse Samples of Urban, Low-Income Preschool Children." *Early Childhood Research Quarterly*, vol. 17, no. 1, 2002, pp. 106–125.
- Fantuzzo, J.W., V.L. Gadsden, and P.A. McDermott. "An Integrated Curriculum to Improve Mathematics, Language, and Literacy for Head Start Children." *American Educational Research Journal*, vol. 48, no. 3, 2011, p.p. 763–793.
- 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, L.S., and S.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, L.S., D. Fuchs, C.L. Hamlett, and P.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, 2008.
- Gao, X., and J. Grisham-Brown. "The Use of Authentic Assessment to Report Accountability Data on Young Children's Language, Literacy and Pre-Math Competency." *International Education Studies*, vol. 4, no. 2, 2011, p. 41.
- 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, R.H., and R.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, C.R., J. Buzhardt, D. Walker, W.J. Howard, and R. 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.
- Greenwood, C.R., J.J. Carta, J. Atwater, H. Goldstein, R. Kaminski, and S. McConnell. "Is a Response to Intervention (RTI) Approach to Preschool Language and Early Literacy Instruction Needed?" *Topics in Early Childhood Special Education*, vol. 33, no. 1, 2013, pp. 48–64.
- Greenwood, C.R., J.J. Carta, J. Atwater, H. Goldstein, R. Kaminski, S. McConnell, and J. Atwater. "The Center for Response to Intervention in Early Childhood: Developing Evidence-Based Tools for a Multi-Tier Approach to Preschool Language and Literacy Instruction." *Journal of Early Intervention*, 36, no. 4, 2015, pp. 1–7.
- Greenwood, C.R., M. Abbott, C. Beecher, J. Atwater, and S.B. Petersen. "Development, Validation, and Evaluation of Literacy 3D: A Package Supporting Tier 1 Preschool Literacy Instruction Implementation and Intervention." *Topics in Early Childhood Special Education*, June 2016, pp. 1–13.

- Grisham-Brown, J., R.A. Hallam, and K. 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.
- Hallam, R.A., A.N. Lyons, K. Pretti-Frontczak, and J. Grisham-Brown. "Comparing Apples and Oranges: The Mismeasurement of Young Children through the Mismatch of Assessment Purpose and the Interpretation of Results." *Topics in Early Childhood Special Education*, vol. 34, no. 2, 2014, pp. 106–115.
- 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, R.L., and K.N. Missall. "Monitoring Preschoolers' Language and Early Literacy Growth and Development." *Young Exceptional Children*, vol. 10, no. 3, 2007, pp. 17–27.
- Hosp, J.L., and S.P. Ardoin. "Assessment for Instructional Planning." *Assessment for Effective Intervention*, vol. 33, no. 2, 2008, pp. 69–77.
- Karelitz, T.M., D.M. Parrish, H. Yamada, and M. Wilson. "Articulating Assessments across Childhood: The Cross-Age Validity of the Desired Results Developmental Profile—Revised." *Educational Assessment*, vol. 15, no. 1, 2010, pp. 1–26.
- Kim, D., and J. Smith. "Evaluation of Two Observational Assessment Systems for Children's Development and Learning." *NHSA Dialog*, vol. 13, no. 4, 2010, pp. 253–267.
- Kim, D., R.G. Lambert, and D.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, R.G., D. Kim, and D.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, S.H., P.R. Swank, J.L. Anthony, and M.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, Y., C. Wang, and S. 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, G.J., and K. Hughes. "Using Individual Growth and Development Indicators to Assess Child and Program Outcomes." *Young Exceptional Children*, vol. 12, no. 1, 2008, pp. 31–41.
- Macy, M.G., D.D. Bricker, and J.K. Squires. "Validity and Reliability of a Curriculum-Based Assessment Approach to Determine Eligibility for Part C Services." *Journal of Early Intervention*, vol. 28, no. 1, 2005, pp. 1–16.
- Macy, M. and S. Bagnato. "The authentic alternative for assessment in early childhood intervention." In *The Oxford Handbook of Child Psychological Assessment*, edited by D. H. Saklofske, C. R Reynolds, and V. L. Schwann. Oxford, UK: Oxford University Press, 2013, pp. 671–681.
- 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, S.R., A.K. Wackerle-Hollman, T.A. Roloff, and M. Rodriguez. "Designing a Measurement Framework for Response to Intervention in Early Childhood Programs." *Journal of Early Intervention*, vol. 36, 2015, pp. 263–280.
- McMann, G., and D. Barnett. "An Analysis of the Construct Validity of Two Measures of Adaptive Behaviour." *Journal of Psychoeducational Assessment*, vol. 2, 1984, pp. 239–247.
- 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.
- Meisels, S.J. "Impact of Instructional Assessment on Elementary Children's Achievement." *Education Policy Analysis Archives*, vol. 11, no. 9, 2003.

- 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 C. Cross, T. Woods, and H. Schweingruber. Washington, DC: National Academies Press, 2009.
- National Survey of Early Care and Education Project Team. "Number and Characteristics of Early Care and Education (ECE) Teachers and Caregivers: Initial Findings from the National Survey of Early Care and Education (NSECE)." OPRE report #2013-38. Washington DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2013.
- Patton, M.Q. *Qualitative Research and Evaluation Methods*, 3rd ed. Thousand Oaks, CA: Sage, 2002.
- Peisner-Feinberg, E., and V. 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, E., and V. 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.
- Piasta, S. "Assessment-Guided Differentiated Instruction to Support Alphabet Knowledge." *The Reading Teacher*, vol. 68, no. 3, 2014, pp. 202-211.
- Pretti-Frontczak, K., S. Bagnato, M. Macy, and D. 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.
- Raudenbush, S.W., and S. Sadoff. "Statistical Inference When Classroom Quality Is Measured with Error." *Journal of Research on Educational Effectiveness*, vol. 1, no. 2, 2008, pp. 138-154.
- Ritchie, J., and L. 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, A.D., S.W. Duggar, L. Moats, M. Glover, and B. 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.
- Sekino, Y., and J. Fantuzzo. "Validity of the Child Observation Record: An Investigation of the Relationship between COR Dimensions and Social-Emotional and Cognitive Outcomes for Head Start Children." *Journal of Psychoeducational Assessment*, vol. 23, no. 3, 2005, pp. 242–260.
- 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, A.M., C. Broussard, M. Fabre, J. Stanley, J. Legendre, and R. 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, A.M., P.A. Snyder, C. Broussard, and K. 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, J., and D.M. Bolt. "Effect of Technology-Enhanced Continuous Progress Monitoring on Math Achievement." *School Psychology Review*, vol. 36, no. 3, 2007, pp. 453–467.
- Zweig, J., Irwin, C. W., Kook, J. F., & Cox, J. (2015). *Data collection and use in early childhood education programs: Evidence from the Northeast Region* (REL 2015–084). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from <http://ies.ed.gov/ncee/edlabs>.

APPENDIX A

EDIT INTERVIEW PROTOCOL

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

EDIT TEACHER INTERVIEW PROTOCOL HIGHLIGHTS

Read Script

- *General Intro Q:* How do you use the information you collect through observations or other ongoing assessments to inform what you do with children to help them learn? Using [FOCAL CHILD] as an example, please walk me through your process for collecting information, reviewing the information, and then using the information to inform instruction.

I. CONDUCTING ASSESSMENTS

I'd like to ask you about how you conduct ongoing assessments.

- Types of information collected about [FOCAL CHILD]. *Refer to video activities and assessment documents.*
- Decide which learning objectives to collect information about (i.e., choose the knowledge, behavior, or skill to be assessed)
- Plan what information to collect, look for opportunities to assess as they present themselves in the day, or both.
- How often collect information about FOCAL CHILD and a specific learning objective. Example.
- Tools used to help assess children (e.g., anecdotal records, checklists, photographs, samples of work, standard tasks, and standardized assessments).
- Assess a particular skill or knowledge in the same way each time or use different tasks or activities. Example.
- Help collecting information from others (e.g., other teachers, parents).
- **IF ANECDOTAL RECORDS PRESENT:** When usually observe and take notes; how much time spent documenting and taking notes.

IF USING A SYSTEM WITH COMPUTER ENTRY: How often enter data [photos, anecdotal records, information from checklists] into the system; when enter data; help entering data (e.g., co-teacher).

IF NOT A COMPUTERIZED SYSTEM: How store and organize the information collected.

II. ORGANIZATION AND INTERPRETATION

Now, I'd like to hear about how you review the information that you collect.

- *Refer to documentation and videos.* How use work samples; How use information from activities like the ones in the videos.
- **IF USING A SYSTEM WITH RATINGS:** Help to evaluate or rate [FOCAL CHILD] (e.g., input from parents, co-teachers); assistance (e.g., a mentor, peer teacher) in reviewing and interpreting ratings (think about the results and what they mean).
- How know if [FOCAL CHILD] needs additional support or a new approach; Use benchmarks or guidelines;

- How decide if [FOCAL CHILDREN] is making the progress that he or she should be making in a given learning objective or domain (between reporting periods AND within reporting periods); how do you determine whether there is a change in the rate of progress and how use that information; how often review the information collected to look at progress; look at progress in a different way when children are doing well like [FOCAL CHILD 1] than when there are some struggles like [FOCAL CHILD 2] or look at progress in the same way for all children.
- How review the information collected (by domain, by objective; one individual child vs. whole class; at one time point or over time). Methods or tools used to review. Examples.
- Assessment information can be used for a lot of different purposes. How use the information (e.g., report to managers, communicate child progress to families – what and how, use for planning lessons or other instructional decisions – what and how to teach)? Examples.

III. LESSON PLANNING AND INSTRUCTIONAL DECISIONS

Now I would like to talk about lesson planning and instruction.

- *Refer to lesson plan/instructional documentation.*
- Frequency of lesson planning.
- Sources of information draw on when planning instruction for individual child (performance AND progress). Example of an instructional decision made that was based on information collected.
- Name of the curriculum. Sources for instructional strategies/approaches/activities.
- Use of adaptations or modifications designed to meet specific child needs (e.g., grouping; peer supports; adult supports; environmental supports).
- Share an example of an individual intervention used with particular children and what were the results.
- Evaluate the success of instructional plans; Method to document and keep track of whether individual interventions were successful for specific children; Collect additional information (re-assess) during or after the individual learning plan; Decide to continue, change, or stop the individual intervention.
- Confer with others; Communicate with co-teacher about how to work with specific children; Guidance on lesson planning (e.g., Ed coordinator; mentor).

IV. SUCCESSES AND CHALLENGES

Read Script

V. WRAP-UP

Read Script

EDIT TEACHER INTERVIEW PROTOCOL

- During the interview, I will be taking some notes about our discussion. To help me keep track of your responses to the questions, with your permission, I will audio-record our conversation. Again, this information will not be shared with anyone from the program; it is meant simply to serve as a record of what you and I discussed. Is that okay? **[INTERVIEWER: Start recording.]**
- An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. The OMB control number for this collection is 0970-0355 and it expires 03/31/2018.
- Our project team is developing a tool called the Examining Data Informing Teaching (the EDIT) to understand how teachers use ongoing assessments. I'd like to ask about your experience using *ongoing assessment*—in other words, how you collect information about children's progress on a regular basis. I'm also interested in hearing about how you use the information you collect.
- This interview will last about an hour. At the very end of our discussion, I will ask you to fill out a brief questionnaire. It takes about 5 minutes. As a reminder, your participation in this project is completely voluntary. All the information you share with me will remain private; no one from your program will see or hear your responses. You and others in your classroom, your students and your center will not be identified by name in any published reports. The information we obtain will be used for research and educational purposes to make the EDIT tool better and to teach researchers how to use the tool. Project documents will not refer to individual people, classrooms, or programs.
- Do you have any questions before we begin the interview?
- To start, I have a few quick questions about your classroom and the focal children that you selected:

How many teachers are in your classroom?

And how many children?

What is the age range of the children in your class?

IF CANNOT INFER FROM CLASS AGE RANGE: And how old is [FOCAL CHILD 1] and [FOCAL CHILD 2]?

How did you choose FOCAL CHILDREN as the children to focus on?

INTERVIEWER: Probe to understand whether the teacher used evidence to select the focal children or based the selection on her own impressions.

What do you see as each focal child's strengths and challenges in language and literacy?

How about social skills?

INTERVIEWER: If the teacher's documentation did not include documentation of the video recorded activities, ask if the teacher has any documentation available. If so, review the documentation with the teacher.

I. CONDUCTING ASSESSMENTS

- First, I'd like to ask you about how you conduct ongoing assessments.

How do you use the information you collect through observations or other ongoing assessments to inform what you do with children to help them learn? Using [ONE OF THE TWO FOCAL CHILDREN] as an example, please walk me through your process for collecting information, reviewing the information, and then using the information to inform instruction.

INTERVIEWER: See if the teacher will give an overview of the whole process in response to this initial question. Then use the remaining questions to probe as needed.

What types of information do you collect about [ONE OF THE TWO FOCAL CHILDREN]?

Do you do activities like the ones you did in the videos? **INTERVIEWER:** You may list the video activities.]

Observations/anecdotal records? **INTERVIEWER:** Use the program's terminology.]

Samples of the child's work or play?

Others (maybe checklists, photos or videos)?

How do you decide which learning objectives to collect information about for [ONE OF THE TWO FOCAL CHILDREN]?

In other words, how do you choose the knowledge, behavior, or skill that you will assess for that child on a given day?

INTERVIEWER: Tailor probes based on documentation.]

- Do you plan what information you will collect or do you look for opportunities to assess as they present themselves in the day or both?
 - ***IF YES:*** Tell me about the planning process. Is it part of lesson planning or a separate process?

How often do you collect information about [ONE OF THE TWO FOCAL CHILDREN]?

How do you decide how often to collect information on a specific learning objective?

- ***IF SOME LEARNING OBJECTIVES HAVE 3 OR MORE PIECES OF EVIDENCE:*** I noticed that you had several [name types of documentation] that provide information about [learning objective]. Can you tell me more about that? How do you decide when to collect more frequent assessment information and what do you do with that information?
- ***IF ONLY 1-2 PIECES OF EVIDENCE FOR EACH OBJECTIVE, ASK:*** Are there some learning objectives that you collect information about more frequently? How do you decide when to do this?

- Do you collect information with the same frequency for all children?

Do you assess a skill or learning objective in the same way each time, or do you use different activities?

- ***IF DIFFERENT ACTIVITIES:*** Can you provide an example and explain why you used different activities?

- **IF SAME WAY:** Can you provide an example and talk about how you decided to use that activity.

How do you know if the child you are working with understands what you are asking them to do during an assessment activity?

Does anyone help you collect information about [ONE OF THE TWO FOCAL CHILDREN], such as an assistant teacher?

Do parents or others at home help you collect information about [ONE OF THE TWO FOCAL CHILDREN]?

Do you take the same approach to collecting information about other children in the classroom?

Do you collect information with the same frequency for each child?

Do you collect information on the same learning objectives for each child?

- Do you do the same types of activities with each child?

Are there children in the classroom who are dual language learners or who have special needs?

IF YES: How do you approach collecting information for those children?

- **IF YES:** For a given activity, like [EXAMPLE FROM VIDEOS], would you do the activity in the same way with each child?

IF NO: Can you give me an example of how you might change an activity?

IF ANECDOTAL RECORDS PRESENT: When do you usually observe and take notes about a child? Do you just observe and take notes or are you interacting with the child at that time (e.g., asking questions, commenting on what they are doing or making suggestions)?

When do you collect most information about what a child knows and can do? Is it during a whole-class learning activity, small-group activities, free choice activities, or do you pull the child aside and work him or her separately?

Do you feel like gathering information about a child with the [ASSESSMENT SYSTEM/METHOD] takes time away from instruction? About how much time do you spend documenting and taking notes each day [if collect information less often, ask “each week”]?

IF USING A SYSTEM WITH COMPUTER ENTRY: Tell me about how you typically enter observations/anecdotal records?

How often do you enter data [photos, anecdotal records, information from checklists] into the system?

When do you enter data into the system?

Does anyone help you enter data into the system, such as assistant teacher?

- Do you record observation data in any other ways, such as taking notes by hand at first or using a checklist and then entering them into a computer?
- **IF NOT A COMPUTERIZED SYSTEM:** How do you store and organize the observations/anecdotal records, checklists, and other information that you collect?

INTERVIEWER: Add probes as needed based on documents and videos.

II. ORGANIZATION AND INTERPRETATION

In this next section, I'd like to hear about how you review the information that you collect.

How do you use work samples (in other words, examples of children work)?

How do you use information from activities like the ones you did in the videos?

INTERVIEWER: Probe about one of the focal children.

How do you decide if [ONE OF THE TWO FOCAL CHILDREN] is making enough progress in a given learning objective or domain?

How do you decide if [ONE OF THE TWO FOCAL CHILDREN] has met a performance goal or developmental expectation?

IF USING A SYSTEM WITH RATINGS: Does anyone help you evaluate or rate [ONE OF THE TWO FOCAL CHILDREN]?

For example, do you rate [ONE OF THE TWO FOCAL CHILDREN] with your assistant?

Do you ask for parent input on ratings for [ONE OF THE TWO FOCAL CHILDREN]?

IF USING A SYSTEM WITH RATINGS: Does anyone assist you in reviewing and interpreting your ratings for [ONE OF THE TWO FOCAL CHILDREN]? Please tell me about how that works.

Do you receive any type of assistance to help you gauge whether your ratings are accurate, such as assistance from leadership staff, a mentor, or a peer teacher?

Does anyone help you think about the results and what they mean (for example, whether [ONE OF THE TWO FOCAL CHILDREN] is making adequate progress)?

How do you know if [ONE OF THE TWO FOCAL CHILDREN] needs additional support or a new approach?

Do you use benchmarks or guidelines?

If so, where do they come from (for example, are they provided by the assessment system or another source)?

How do you decide if [ONE OF THE TWO FOCAL CHILDREN] is making the progress that he or she should be making in a given learning objective or domain?

Do you look at progress in a different way when children are doing well like [FOCAL CHILD 1] than when there are some struggles like [FOCAL CHILD 2], or do you look at progress in the same way for all children?

How do you decide if the progress that they are making between reporting periods is enough for them to be where they need to be in those skills by the end of the year?

How often do you review the information that you collect to see [ONE OF THE TWO FOCAL CHILDREN]'s progress over time?

Do you ever look at the change in [ONE OF THE TWO FOCAL CHILDREN]'s progress within a reporting period, in addition to looking at the change between reporting periods?

How do you review the information that you collect?

Do you look at how [ONE OF THE TWO FOCAL CHILDREN] is doing in each learning domain and objective?

- For example, do you compare how [ONE OF THE TWO FOCAL CHILDREN] is doing in different domains, such as language, literacy, or social-emotional development?
- If so, do you look at the change in rate of [ONE OF THE TWO FOCAL CHILDREN]'s progress over time?
- Do you compare [ONE OF THE TWO FOCAL CHILDREN]'s progress to the progress of other children in the class? If so, how?
- Do you look at how the entire class is doing in each learning domain? Do you look by objective? For example, do you look at a particular domain, like literacy, and see how all children are doing?
- If so, do you look at the change in progress for the entire class over time?
- Do you look at groups of children (for example all the 4 year olds)?

What do you do if you see inconsistency in [ONE OF THE TWO FOCAL CHILDREN]'s performance on the same task over time?

- For example, if [ONE OF THE TWO FOCAL CHILDREN] does well on a task one day and poorly on that task the next day, how do you explain that?

We've been talking about what you do with [FOCAL CHILDREN]. Now let's think about the other children in the class. Is there anything else you do, or anything you do differently, when you review the information you collect for the other children in the class?

Assessment information can be used for a lot of different purposes. How do you use the information from the [ASSESSMENT] (e.g., report to managers, communicate child progress to families, use for planning lessons or other instructional decisions)?

Do you share information from the assessments with families?

- **IF YES:** How do you decide what you share?
- How do you share it with families?

INTERVIEWER: Add probes as needed based on documents and videos.

III. LESSON PLANNING AND INSTRUCTIONAL DECISIONS

Now I would like to talk about planning for instruction. We use the word ‘instruction’ to include all that you do to help children learn and grow, including center activities that you provide, small group interactions, books, songs, questions, and other interactions that you have with children. Some instruction may be an “in the moment,” but most teachers plan at least some of their instruction.

How do you plan your instruction? **[INTERVIEWER: Listen for daily, weekly, monthly.]**

What information do you use when planning? **[INTERVIEWER: Many teachers may select a unit aligned with a season and planning may be divorced from assessment; or teachers may depend only on child interest in a topic. They are trusting the curriculum to ‘cover’ all the skills needed.]**

When you are planning what you will do for the next [FILL IN TEACHER TIME PERIOD FOR PLANNING] do you also plan to collect specific information about children? How do you decide what you will collect?

Do you use any of the information that you collect about children to inform your planning, such as decisions about what to teach, which activities or units to use, or how to support children’s learning?

IF YES: Please give me an example of how you do this. Walk me through how you use the information in planning instruction or making instructional decisions.

- Do you consider just current performance or do you also consider how much progress children are making in a particular area?
- You mentioned that you draw on [INSERT EVIDENCE TEACHER USES FOR ASSESSMENT] to make decisions about child progress. How does it affect your decisions about instruction and about what to assess/observe?
- Please give an example of an instructional decision that you made that was based on information that you collected about a child. **[INTERVIEWER: Prompt from documentation if available.]**

INTERVIEWER: Listen for at the individual, small group, and/or whole class level.

What ideas or sources do you use in selecting instructional strategies or approaches?

- Instructional approaches could be additional practice, how you group children, prompting or questioning or using pictures or objects to help children understand something, etc.

INTERVIEWER: If not mentioned, probe about use of adaptations or modifications designed to meet specific child needs; peer supports; adult supports; and environmental supports.

How do you communicate with other team members, like your assistant, on how to work with specific children?

Does anyone give you input or guidance on how to use the information you collect to guide lesson planning, such as a mentor?

FOR EACH TYPE OF PLANNING DOCUMENT INCLUDED IN THE DOCUMENTATION:

How do you use the [DOCUMENT NAME]?

Did you personally create the [DOCUMENT NAME]?

- If so, how did you create it?
- What information did you draw on to create it?
- How often do you create or revise the [DOCUMENT NAME]?

For which children do you create the [DOCUMENT NAME]?

- How do you use the [DOCUMENT NAME]?

Do all children experience all the activities?

IF NO: How do you decide which activities or lessons children will get?

How do you group children?

IF THE TEACHER USES SMALL GROUPS: How do you determine which children are in which groups?

Do you plan additional activities for individual children?

IF YES: What types of strategies do you use in an individual activity?

Do you ever decide to have a child do something differently from others in the class, for example, use different materials or only do parts of an activity or do more of something? Do you individualize instruction within a group by using any of the following?

INTERVIEWER: If the teacher mentions any of the following, ask for a description/example.

- Plan different types of questions for different children?
- Use cues or different prompts for some children?
- More opportunities for practice?
- Change something in the environment? (e.g. changing seating or positioning, going to a quieter area, providing visual prompts, etc.)
- Use peers to help a child or model for a child (Peer strategies)?
- **[Other adaptations and modifications]** Are there any other changes that you make to an activity to support an individual child's learning?
- **[If teacher mentions child interests or strengths]** How do you incorporate a child's interests or strengths into instruction?

Are you ever in a situation where you want to make one of these modifications for a child but are unable to? **[INTERVIEWER: Note that we are trying to get at if the teacher faced any logistical constraints.]**

Do you have a method for recording whether your instructional plan was implemented?

How do you evaluate the success of your instructional plans?

- How do you know whether your small groups or individualized learning plans work?
- Do you have a method to document and keep track of whether individual interventions were successful for specific children?

- How do you decide to continue, change, or stop the individual intervention?
- [IF NOT DISCUSSED ALREADY:] Do you collect additional information (re-assess or conduct additional observations) during or after the individual learning plan? What do you do with that information? Please give an example.
- Can you share an example of an individual intervention that you have used with particular children and what were the results?

INTERVIEWER: Add probes as needed based on documents and videos.

IV. SUCCESSES AND CHALLENGES

Finally, I would like you to reflect on your experience using assessments to individualize instruction in your classroom.

What would you consider successes? In other words, what are the most useful aspects of your use of ongoing assessment from your perspective?

What factors or circumstances contributed to these successes?

What challenges have you experienced with using ongoing assessment information?

- Are there any parts of the assessment process in your program that you find are not particularly useful? If so, tell me more.

What did you do to address these challenges, and how successful has that been?

V. WRAP-UP

Is there anything else we haven't covered that you'd like us to know about using assessments in your classroom?

REMIND TEACHER TO COMPLETE SHORT SAQ

PROVIDE GIFT CARD AND FILL OUT RECEIPT

- **LET TEACHER KNOW THAT DEBRIEF IS COMING SOON—ASK FOR TIMES THAT ARE GENERALLY GOOD**

- Thanks again for your time and your valuable insights!

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

APPENDIX B

TEACHER SELF-ADMINISTERED QUESTIONNAIRE (SAQ)

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

Examining Data Informing Teaching (EDIT)

Teacher Questionnaire

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 private 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. It should take about 5 minutes to complete. 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

3

Thank you very much for your help.

An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. The OMB control number for this collection is 0970-0355 and it expires 03/31/2018.

A. STAFF DEVELOPMENT	B. COMMUNICATION WITH PARENTS
<p>A1. Please record today's date:</p> <p> _ _ / _ _ / 2 0 _ _ </p> <p>A2. How many hours a year do you attend staff trainings about assessment or evidence-based instructional practices?</p> <p> _ _ _ HOURS</p> <p>A3. How often do you have one-on-one supervision meetings or group supervision meetings about assessment or evidence-based instructional practices?</p> <p>MARK ONE ONLY</p> <p>0 <input type="checkbox"/> Never</p> <p>1 <input type="checkbox"/> Once a year</p> <p>2 <input type="checkbox"/> A few times a year</p> <p>3 <input type="checkbox"/> Once every 2 months</p> <p>4 <input type="checkbox"/> Once a month</p> <p>5 <input type="checkbox"/> Two times per month</p> <p>6 <input type="checkbox"/> One time per week</p> <p>7 <input type="checkbox"/> More than once a week</p> <p>n/a <input type="checkbox"/> Not applicable</p> <p>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?</p> <p>1 <input type="checkbox"/> Yes</p> <p>0 <input type="checkbox"/> No</p>	<p>B1. How often do you talk to parents about how their children are doing on a formal or informal basis?</p> <p>MARK ONE ONLY</p> <p>0 <input type="checkbox"/> Never</p> <p>1 <input type="checkbox"/> Only at parent-teacher conferences</p> <p>2 <input type="checkbox"/> Every 2 or 3 months</p> <p>3 <input type="checkbox"/> Once or twice a month</p> <p>4 <input type="checkbox"/> Once or twice a week</p> <p>5 <input type="checkbox"/> Daily</p> <p>B2. How often do you hold formal parent-teacher conferences with parents about individual children?</p> <p>MARK ONE ONLY</p> <p>0 <input type="checkbox"/> Never</p> <p>1 <input type="checkbox"/> Once a year</p> <p>2 <input type="checkbox"/> Twice a year</p> <p>3 <input type="checkbox"/> 3 times a year</p> <p>4 <input type="checkbox"/> 4 or more time a year</p>

C. EMPLOYMENT	D. EDUCATION AND EXPERIENCE
<p>C1. Are you currently working in this early childhood setting full or part-time? MARK ONE ONLY</p> <p>1 <input type="checkbox"/> Full time</p> <p>0 <input type="checkbox"/> Part time</p> <p>C2. Counting this school year, how long have you worked in your <u>current early childhood setting</u>?</p> <p> _ _ YEARS _ _ MONTHS</p> <p>C3. Counting this school year, how long have you worked in your <u>current classroom</u>?</p> <p> _ _ YEARS _ _ MONTHS</p> <p>C4. Please indicate your role(s) at this early childhood setting. MARK ALL THAT APPLY</p> <p>1 <input type="checkbox"/> Owner</p> <p>2 <input type="checkbox"/> Director</p> <p>3 <input type="checkbox"/> Lead Teacher</p> <p>4 <input type="checkbox"/> Assistant Teacher</p> <p>5 <input type="checkbox"/> Teacher</p> <p>6 <input type="checkbox"/> Administrative Assistant</p> <p>7 <input type="checkbox"/> Other role (<i>please specify</i>)</p> <p>_____</p>	<p>D1. Do you currently hold a Child Development Associate (CDA) credential?</p> <p>1 <input type="checkbox"/> Yes</p> <p>0 <input type="checkbox"/> No</p> <p>D2. What is the highest level of education you have completed? MARK ONE ONLY</p> <p>0 <input type="checkbox"/> High school diploma or GED</p> <p>1 <input type="checkbox"/> College course(s) without a degree</p> <p>2 <input type="checkbox"/> Associate's degree</p> <p>3 <input type="checkbox"/> Bachelor's degree</p> <p>4 <input type="checkbox"/> Master's degree</p> <p>5 <input type="checkbox"/> Education specialist or professional diploma based on at least one year of course work past a Master's degree level</p> <p>6 <input type="checkbox"/> Doctorate</p> <p>7 <input type="checkbox"/> Other (<i>please specify</i>)</p> <p>_____</p> <p>D3. In what field did you obtain your highest degree? MARK ALL THAT APPLY</p> <p>1 <input type="checkbox"/> Child development or developmental psychology</p> <p>2 <input type="checkbox"/> Early childhood education</p> <p>3 <input type="checkbox"/> Elementary education</p> <p>4 <input type="checkbox"/> Special education</p> <p>5 <input type="checkbox"/> Other (<i>please specify</i>)</p> <p>_____</p> <p>D4. How many college courses have you completed related to child development and/or assessment?</p> <p> _ _ CHILD DEVELOPMENT COURSES</p> <p> _ _ ASSESSMENT COURSES</p> <p>D5. Including this year, how many years have you worked with preschool aged children?</p> <p> _ _ YEARS</p>

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 [STAFF], at [PHONE NUMBER].

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 C

LETTER TO TEACHERS

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




[NAME]
EDIT Project Director

[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 [VISIT DATE]. To help you get your materials collected and organized for our visit, we ask that you prepare as follows:

<p>First a phone call [CONTACT]</p> 	<p>On this brief call, we will:</p> <ul style="list-style-type: none"> • Review procedures to help you prepare for our visit • Provide technical assistance • Answer any questions you have about the project
<p>Next</p>  <p>Select 2 Focal Children</p>	<p>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.</p> <p>CONSIDER SELECTING: [INSERT INITIALS OF CONSENTED CHILDREN]</p> <p>DO NOT CONSIDER: [INSERT INITIALS OF NON-CONSENTED CHILDREN]</p>
<p>Week of [DATE]</p>  <p>Collect Round 1 Videos</p> <p>*Please only include children who have permission to participate in recordings.</p>	<p>Four separate videos*</p> <p><u>Child doing well (in a small group or individually):</u></p> <ol style="list-style-type: none"> 1. Working on language, literacy or social skills activity/lesson 2. Activity where you are collecting information about the child's knowledge or social skills <p><u>Child with challenges (in a small group or individually):</u></p> <ol style="list-style-type: none"> 3. Working on language, literacy, or social skills activity/lesson 4. Activity where you are collecting information about the child's knowledge or social skills

LETTER TO: [TEACHER NAME]
 FROM: [PROJECT DIRECTOR]
 DATE: [DATE]
 PAGE:




<p>Week of [DATE]</p>  <p>Collect Round 2 Videos</p> <p>*Please only include children who have permission to participate in recordings.</p>	<p>Two separate videos*</p> <p><u>Child doing well:</u></p> <ol style="list-style-type: none"> Working on a language, literacy, or social skills activity/lesson <p><u>Child with challenges:</u></p> <ol style="list-style-type: none"> Working on a language, literacy, or social skills activity/lesson
<p>Between Today and [DATE]</p>  <p>Gather Documents</p>	<p>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.</p> <p>Examples of what some teachers use:</p> <ul style="list-style-type: none"> Anecdotal records, photos, checklists, charts, graphs Portfolios of children’s work Family reports, class reports Lesson plans; individualized learning plans Instructional sequence, unit plans, yearly themes or projects, schedules for observing or collecting information about children
<p>[DATE]</p>  <p>EDIT Team Visit</p> <p><i>Please note: We are not observing your classroom <u>in person</u> during our visit, so the focal children you selected do not have to be present in school on the day of our visit.</i></p>	<p>On [DATE], 2 researchers will visit your center to:</p> <ul style="list-style-type: none"> Review documents you gathered Review the videos you recorded Speak with you for one hour to learn more about how you use assessment data to plan strategies for children in your classroom Ask you to complete a brief paper questionnaire about your educational and teaching experiences

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 center with a \$50 gift card for purchasing materials for the center in appreciation for your participation in this important study.

LETTER TO: [TEACHER NAME]
 FROM: [PROJECT DIRECTOR]
 DATE: [DATE]
 PAGE:

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

Preparation	Tip	Note
<p>Select 2 Focal Children</p> 	<p>Think about the children in your class and select one child performing well in language and literacy and another child who is struggling in language, literacy, or social skills.</p>	<p>Please do not select children who do not have permission to participate.</p> <p>CONSIDER SELECTING: [INSERT INITIALS OF CONSENTED CHILDREN]</p> <p>DO NOT CONSIDER: [INSERT INITIALS OF NON-CONSENTED CHILDREN]</p>
<p>Videos with Focal Children</p> 	<p>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.</p> <p>For each of the videos, please focus on language or literacy learning or social skills, but otherwise follow your typical classroom practice.</p> <p>Only collect information in the ways that you typically do in your classroom.</p> <p>Confirm that your videos saved to the iPad after each recording session.</p>	<p>Do not include other non-consented adults or non-consented children – even in the background – in your videos.</p>
<p>Gather Documents</p> 	<p>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.</p>	<p>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.</p> <p>We will return all documentation at the end of our visit; names will be hidden in any copies/photographs made of the documentation.</p>

If you have any questions, please contact me by email at [PROJECT DIRECTOR EMAIL] or by phone at [PROJECT DIRECTOR PHONE]. Thank you in advance for your time and consideration.

Sincerely,

This collection of information is voluntary. Public reporting burden for this collection of information is estimated to average 205 minutes per teacher, including the time for reviewing instructions, gathering and maintaining the data needed, and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. The OMB control number for this collection is 0970-0355 and it expires 03/31/2018.

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

APPENDIX D
TRAINING PLAN

This page has been left blank for double-sided copying

APPENDIX D. TRAINING PLAN

To extend development of the EDIT beyond the pretest stages, a training session would be necessary to proceed with further 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.

Draft training materials could include an agenda (Appendix F), manual (drawing from Appendix E), presentations, training exercises, and a certification plan.

The EDIT study team provides an item-by-item description of the instrument with examples and definitions of terms developed based on our pretest experiences for inclusion in a training manual (see Appendix E). The trainers should review contents of the manual with trainees during training. Ideally, the observation training would 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 should include multiple opportunities to practice with the measure by rating documents, watching and rating videos, and listening to recordings of teacher interviews, followed by group discussion.

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

- **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**
- **Proficiency in administration of the semi-structured teacher interview**
- **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 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 setting staff and teachers at each center**

We recommend that the training format 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 to learn how to complete the EDIT.** Discussions between trainers and trainees to justify scores to particular items could follow. Trainers can 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 should apply the rubrics to live observations in local classrooms. Trainers could accompany trainees to each classroom during the field practice. Following the practice observation, the trainees should 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 should culminate in a certification process designed to reflect field procedures for the site visit and EDIT rating processes.³⁹ Trainees would be required to rate using the EDIT rubrics and a standard set of materials designed to reflect EDIT procedures. The standard set of materials could include mock documents, as well as videos and recordings or transcripts of interviews. After certification, a trainee will fall into one of three categories:

1. **Fully certified.** Trainees who pass certification would be permitted to conduct EDIT site visits.
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 training team. Provisionally certified trainees would 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 would not be certified to administer the EDIT.

³⁹ Reliability thresholds on the EDIT should be determined in consultation with ACF.

APPENDIX E

TRAINING DEFINITIONS AND EXAMPLES

This page has been left blank for double-sided copying

Overarching definitions

- *Ongoing child assessment*
 - Ongoing child assessment entails repeated assessments and observations of a child's performance and progress over time.
- *Individualization*
 - Individualization is a process in which a teacher uses data to identify a child's skill level for a learning goal and tailors instruction for that child. The teacher uses data on an ongoing basis to see whether the child is progressing in response to the instructional changes and adjusts instruction as needed.
- *Performance*
 - Throughout this instrument, references to "performance" are meant to indicate a child's score on an ongoing assessment as measured at a single time-point. This is differentiated from "progress" which is measured over time.
- *Progress*
 - Throughout this instrument, references to "progress" are meant to indicate changes in performance *within a reporting period*. To assess progress, a teacher must have at least 3 data points within a reporting period.
- *Reporting Period*
 - Throughout this instrument, references to "reporting period" are meant to indicate the times of year at which a program requires its teachers to document and review progress for each child in their classrooms. While the timing of reporting periods vary, they typically occur two to three times per year.

⁴⁰ Many of the examples provided in this manual are drawn from the Learning from Assessment materials (Atkins-Burnett et al. 2014) provided on the Head Start Early Childhood Learning and Knowledge Center.

Cover page

Teacher: _ _ _ _ _ _ _ _ Date: _ _ _ / _ _ _ / _ _ _ _ _ _ _ _	Rater: _ _ _ _ _ _ _ _
Focal Child 1 ⁴¹ : _ _ _ _ _ _ _ _ Focal Child 2: _ _ _ _ _ _ _ _	Document review: Start time: _ _ _ : _ _ _ AM/PM End time: _ _ _ : _ _ _ AM/PM
List of types of artifacts received (Place a “(B)” in front of artifacts if they were collected more than 2 months ago and are <i>background</i> to current performance and progress.) _____ _____ _____	

Orientation to the EDIT’s multiple data sources

The EDIT is a multi-method measure which consists of a document review, video-based observations, and a one-hour teacher interview with a reflective think aloud protocol. Video-based observations will be recorded over a two- to three-week period and followed by a one-day visit to conduct the document review, rate the videos, and conduct the teacher interview.

For the document review, the teacher provides ongoing assessment data for two students, one performing well and the other facing challenges. The assessor reviews current lesson plans for evidence of individualization and then rates the documents with rubrics, checklists, and rating.

For the video-recorded classroom observations, the teacher first video-records a combination of assessments and small-group instruction that includes one or both of the focal children. The assessor views video during visit after rating the documents and evaluates the observations using rubrics, checklists, and ratings.

⁴¹ Focal child 1 always refers to the child performing well. Focal child 2 always refers to the child experiencing challenges.

At the end of the visit day, the assessor conducts a one-hour teacher interview. The assessor probes for additional explanations about the artifacts and video data and revises prior ratings to reflect new information learned during the interview.

Orientation to this training guide

In the pages that follow, we provide definitions and examples for each section of the EDIT. For holistic and analytic rubrics, items in red text are those for which we provide an explicit definition and/or example; items in black text are (1) those that do not require clarification or (2) items closely tied to items in red text that do not require clarification beyond the information given for the items in red text.

Definitions and examples for selecting the assessment target

1. HOLISTIC RUBRIC FOR “SELECTING THE ASSESSMENT TARGET”

Assessment targets are linked to meaningful outcomes (that is, a skill, knowledge, or behavior that a child needs to be successful, now or in the future). The targeted behavior is developmentally appropriate (for example, (1) looking at combinations of 5 with manipulatives rather than asking children to solve written equations, or (2) asking children to identify which spoken word is unlike the others in a series of 5 words rather than to compose their own rhyming poem). To meet Head Start and other early childhood education policy requirements, the targets include assessment of key domains related to school-readiness: language and literacy development, cognition and general knowledge, approaches to learning, physical well-being and motor development, and social and emotional development. The targets are generalizable in that the skill, knowledge, or behavior can be demonstrated across settings. The targeted skills or behaviors are those in which children of this age typically make progress within the program year (that is, change is expected in the current year). The targets (1) address areas that are taught in the classroom curriculum, and (2) are defined and measured based on observable behaviors. Targets can be assessed universally (that is, for all children), or evidence of individualization is clear with appropriate accommodations made as necessary for specific children to demonstrate progress.

Sources: Documents (end-of-year goals, assessment objectives or protocols, curriculum/instructional sequence), observations, teacher interview.

Assessment target

- The assessment target is the knowledge, skill, or behavior that the teacher wants to assess.
- Examples of assessment targets in preschool:
 - Recognize shapes or colors when they are name
 - Show understanding of cause-and-effect relationships
 - Follow a two-step direction
 - Persist in assembling a puzzle with fewer than 20 pieces
 - Take turns with another child when playing a matching game

1	3	5	7	ENTER RATING
<p>Targets not clearly linked to structured curriculum or meaningful outcomes.</p> <p>OR</p> <p>Targets not developmentally appropriate.⁴²</p>	<p>Defines at least one target that is linked to structured curriculum and meaningful outcomes.</p> <p>At least one target is defined for one of the five key school-readiness domains.</p> <p>At least one target is measured using observable behaviors.</p> <p>Targets may not be in an area in which children typically make progress within the program year (that is, cannot measure change over time).</p>	<p>Defines a few targets that are linked to structured curriculum meaningful outcomes.</p> <p>A few targets are individually appropriate for the child.</p> <p>At least one target is defined for each of the five key domains for school-readiness per child.</p> <p>Targets are measured using observable behaviors.</p> <p>At least one target per child is measured over time to track progress.</p> <p>Targets are in an area in which children typically make progress within the program year with instruction or intervention (that is, can measure change over time).</p> <p>At least one assessment target is represented in individualized plans for instruction for at least some of the children.</p>	<p>Defines multiple targets that are linked to structured curriculum and meaningful outcomes.</p> <p>Multiple targets are individually appropriate for the child.</p> <p>Multiple targets are defined for three of the five key domains for school-readiness per child.</p> <p>Targets are measured using observable behaviors AND, as appropriate, teacher examines whether targets generalize across settings.</p> <p>At least one assessment target is represented in individualized plans for instruction for all children.</p>	<p>RATING: _____</p>

- *Targets not developmentally appropriate.*
 - The assessment target should be developmentally appropriate. For example, an assessment target in number composition could ask preschoolers to look at different ways to make a set of five using concrete objects—two pegs and three pegs, or four pegs and one peg—but should not ask them to solve written equations without any supports or

⁴² Red text indicates that an example is provided to illustrate that criterion.

context (for example, $2+3= \underline{\quad}$). The teacher should always be working with the child on the next step for the child's individual developmental progress. For example, a child should not be asked to solve written equations without first having a strong understanding of how numbers can be put together. Additionally, when a child has a clear understanding of a concept, the teacher should move to a new concept.

- *Defines at least one target that is linked to structured curriculum and meaningful outcomes.*
 - The target should align with areas taught in the classroom curriculum. If teachers expect children to make progress in a particular area, they should be giving them opportunities to learn that skill, knowledge, or behavior. The curriculum is the road map to ensure that children have those learning opportunities.
 - A **meaningful outcome** is a skill, knowledge, or behavior that a child needs to be successful, either now or in the future.
 - Example of a target linked to a curriculum and meaningful outcome:
 - Counts five objects with one-to-one correspondence
 - Identifies feelings
 - Identifies rhymes in a finger-play or song
 - Writes name
- *At least one target is defined for one of the five key school-readiness domains.*
 - To meet the requirements of Head Start and other early childhood education policies, assessment targets should be in a key **domain related to school readiness**:
 - Language and literacy development
 - Cognition and general knowledge
 - Approaches to learning
 - Physical well-being and motor development
 - Social and emotional development
 - Information about the Head Start Early Learning Outcomes Framework 2015 can be found here: <https://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach/elof>
- *At least one target is measured using observable behaviors.*
 - Teachers should select an assessment target that can be defined and measured based on observable behaviors.
 - Behaviors may be observed through direct assessment. For example, a teacher could assess counting by asking a child how many items are in a set.
 - Behaviors also may be observed through indirect evidence.
 - For example, a teacher could assess a child's approach to learning by observing how long he or she persists at a task.
 - Or a teacher could assess fine motor proficiency by observing a child zipping his or her coat independently.
 - Indirect evidence may also include looking at something the child completed or created. For example, the teacher could assess a child's understanding of

patterns by looking at a sequence of colored blocks that the child created. Note that if the teacher did not see the child complete the task, the teacher may not know whether the child did it independently or with another child, so the teacher's hypotheses or inferences about the child may not be as strong.

- *A few targets are individually appropriate for the child.*
- Targets should be focused on the child's current skills and abilities, rather than their age or what the curriculum dictates. If a child is not yet combining words, it is not realistic to expect that the child would retell a story with a beginning, middle, and end. Similarly, if a child is already very accomplished in an area – performing beyond age expectations – unless the child shows a continuing interest in that area, it would be more appropriate to broaden skill rather than continue to focus on that. For example, if a child is rote counting to 25, it would be more appropriate to work in understanding how to compose (there are 3 chairs at this table. I am going to move one more over to the table. How many children can sit at this table now?) and decompose numbers less than five (I have 4 apples and eat one how many are left?).
- At least 3 targets should be an area that can be changed with instruction or intervention, rather than an area that changes only with maturation. For example, physical height is not amenable to instruction, but vocabulary is something that can change with intervention. If the teacher targets climbing the ladder to the slide when the child's legs are not long enough to take the next step, he or she is targeting an area that cannot be changed with intervention. The key is for the teacher to clearly be showing evidence that the teacher is considering the individual child's developmental progress in an area that can change with intervention. Children who easily name pictures and are showing interest in the letters in their own name might have a target of naming letters.
- *Targets are in an area in which children typically make progress within the program year with instruction or intervention (that is, can measure change over time).*
- Assessment targets should help the teacher understand the child's range of skill, knowledge, or behavior. Children who have poor strength in pincer grasp may have a goal focused on improving finger strength and assessment target of pinching open clothespins or putting together Lego-type blocks rather than writing with a pencil. Working directly on writing with a pencil could lead to poor habits in holding a pencil and would be hard to break later. Alternatively, the assessment target might be to trace their name with a finger or draw letters in sand or finger paint.
- The target should be sensitive to change—that is, it should be an area where children of this age typically change or make progress within the program year. For example, over the course of a year, a child will increasingly come to understand stories, count to higher numbers, and engage in more sophisticated play with peers. Progress due to maturation would not be included here (e.g., the child's physical height changes between reporting periods).

- *At least one assessment target is represented in individualized plans for instruction for at least some of the children.*
- Individualized plans for instruction should clearly link lessons and activities to assessment targets. In the example below, the teacher does this by labelling each of the planned activities with a corresponding objective.

Individualized Learning Plan

Child Name	Domain	Objective	Activity
Grace	Language & Literacy	3b	Play matching game with alphabet cards.
Sam	Social/Emotional Development	9c	Invite child to engage in conversation during small group.
Jose	Math	15a	Solve story problems with addition and subtraction.

- *At least one target per child is measured over time to track change.*
- Examples of targets indicating desired progress:
 - Child traces her name in the fall, writes the first letter of her name in the winter, and writes her full name from memory in the spring.
 - Within a reporting period, the child increases the number of consonant sounds that he can make for letters.
- *Teacher examines whether targets generalize across settings.*
- For a target to **generalize** across settings, the teacher must measure whether the child can meet the target in a variety of contexts, formats, or areas of learning. For example, can the child name a letter when they see it in their name, and when it is on an index card by itself, and when it is at the beginning or end of a word in a book?

Definitions and examples for selecting and implementing the assessment method

2. ANALYTIC RUBRIC AND RATINGS FOR “SELECTING AND IMPLEMENTING THE ASSESSMENT METHOD”
<p>Observation or Assessment Method. Teachers decide how they will gather information about the assessment target (learning objective or behaviors). The method of data collection should focus on observable behaviors, be a fair measure of what children know and can do (including linguistic and cultural appropriateness), and offer the opportunity to easily collect information frequently enough that comparisons across time can be made. Over time, the assessment method should provide information about whether the child generalizes the skills, knowledge, and behaviors and so should provide information from more than one context. The frequency of data collection supports reliable and valid interpretation of child progress.</p> <p><i>Sources: Documents (assessment objectives or protocols, assessment manual, assessment schedules and plans, description of assessment), observations, interview.</i></p>

- *Observation or Assessment method*
 - The **observation or assessment method** is the way that the teacher gathers information about the skill, knowledge, or behavior of interest.
 - In curriculum-embedded approaches, the teacher systematically records developmental observations over time about a child’s naturally occurring behaviors and competencies during daily activities. The field sometimes refers to these assessments as **authentic**.
 - Examples of assessment methods:
 - Observations of a child going about his or her day-to-day tasks and routines
 - Video and audio recordings and photographs
 - Samples of a child’s work or play, such as drawings, dictation of what a child says, and pictures of three-dimensional structures that a child builds
 - Structured tasks, such as asking a child to name pictures, shapes, numbers, or letters using flashcards; zip a zipper; or copy a block structure
 - Standardized probes and questions, such as asking a child, “When is your birthday?” or “Where should I begin reading?”

	1	3	5	7	ENTER RATING
A	No evidence available OR The evidence collected is not aligned with assessment target(s).	The evidence collected is occasionally aligned with the assessment target(s). OR Evidence is very limited but is aligned.	The evidence collected is sometimes aligned with the assessment target(s).	The observed evidence collected is always aligned with the assessment target(s).	RATING: _____

The evidence collected is always aligned with the assessment target(s).

- Teachers should select an assessment method that focuses on the skill, knowledge, or behavior of interest.
- For example, if the target is addressing the child’s persistence, observing only familiar tasks that the child finds easy would not provide a good measure of persistence. Observing a child in a challenging task would provide a stronger measure of persistence.
- To receive a 7, evidence must be observed. The rater can only score based on the evidence we receive in the documentation.

	1	3	5	7	ENTER RATING
B	Assessments not implemented frequently enough to examine child progress.	Only 1 or 2 assessment targets are assessed more than once per child to examine progress.	3 or more assessment targets are assessed more than once per child, and frequency of assessments is usually greater when instruction is more intensive.	Assessment information is collected frequently enough to examine child progress (at least three times per reporting period) or change over time. Frequency of data collection is usually greater when instruction is more intensive.	RATING: _____

- *Assessment information is collected frequently enough to examine child progress (at least three times per reporting period) or change over time.*
 - The schedule for data collection should match the pace of a child’s expected progress over time, enabling the teacher to examine a child’s progress over time in each domain. Specifically, data should be collected often enough that the teacher knows when the child needs more or less support or challenge. For example, the schedule should allow the teacher to review progress *within* a quarter, with a minimum of three observations for a given assessment target within a quarter. The teacher could collect some data each week, with more frequent data collection in the areas being taught each day. An assessment plan can facilitate this process to ensure that each child is assessed on a regular basis and that needed information is collected frequently enough to see change (or lack of change).
- *Frequency of data collection is usually greater when instruction is more intensive.*
 - For the frequency of data collection to align with the intensity of instruction, the teacher should assess more frequently on topics currently covered in the instruction and those that are targeted for individualization. For example, if something was taught the previous month (such as responding to “who” questions about a story) and is being reviewed regularly, the teacher may collect information once every few months after the initial instructional period. When providing daily instruction for a specific learning objective

(such as writing their name or naming letters), the teacher may collect data every one to two weeks. If children are having difficulty with a learning objective and the teacher provides additional practice or more intensive one-to-one instruction, the teacher may collect evidence daily or weekly.

- When rating the EDIT, it is often hard to have evidence of the alignment from the documents we received from the teacher. We typically have to rely on the interview to gather evidence for this indicator.

	1	3	5	7	ENTER RATING
C	Assessment tasks do not provide information about different levels of the knowledge, skill, or behavior of interest.	For some targets, the assessment task differentiates between varying levels of quality OR independence over time, e.g., by breaking down the task into steps and/or requiring the child to complete the task in different contexts.	For most targets, the assessment task differentiates between varying levels of quality OR independence over time, e.g., by breaking down the task into steps and/or requiring the child to complete the task in different contexts.	For all targets, the assessment task clearly differentiates between varying levels of quality OR independence over time, e.g., by breaking down the task into steps and/or requiring the child to complete the task in different contexts.	RATING: _____

- *The assessment task clearly differentiates between varying levels of quality OR independence over time, e.g., by breaking down the task into steps and/or requiring the child to complete the task in different contexts.*
- The assessment method should clearly differentiate between varying levels of a child’s skill or independence over time. In other words, the child should be able to demonstrate progress over time, and the documentation should capture that change.
- To do this, the teacher may **break the task into steps**, documenting which steps the child completes. For example:
 - If the target is learning the entire alphabet, the initial assessment may ask the child to identify five letters; each subsequent assessment could add additional letters.
 - If the target is learning to rhyme, the initial assessment could ask the child to identify which word is unlike the others in a series of five words (with four words that rhyme and one word that does not, such as jump, gump, lump, rat, pump); the later assessment could ask the child to do the same for a series of three words (sent, tent, mop); and the final assessment could ask the child to produce their own rhyme or ask the child to pick out the word that rhymes with “tent” in a series of words with the same initial or final consonant (rake, rat, rent, roll).

- The teacher may also observe and document whether a child can *generalize* the target skill. In other words, can the child apply a skill in different contexts? For example, if the target is learning letters, the child may first identify letters on cards and then in books.
- The teacher may examine differences in the quality of the child’s skills.
 - Can the child walk across the room only with her arms raised for balance, or can she walk with arms down or carrying something?
 - Can the child control the speed of walking (as opposed to falling forward)?
 - How well does the child copy or write her name? Are the letters in the correct order? Turned the correct direction? Connected the way they should be? What is the relative size of the letters?

	1	3	5	7	ENTER RATING
D	All assessments take time away from instruction and practice (>10 minutes).	Assessments typically take time away from instruction and practice (>10 minutes).	Some assessments occur in the context of instructional activities OR in a way that allows teacher to efficiently obtain information about child’s knowledge/skill/behavior in a clear task (such as naming pictures, letters, or shapes).	Assessments typically occur in the context of instructional activities OR in a way that allows teacher to efficiently obtain information about child’s knowledge/skill/behavior in a clear task (such as naming pictures, letters, or shapes).	RATING: _____ <input type="checkbox"/> CANNOT RATE

Assessments typically occur in the context of instructional activities.

- Teachers should conduct the assessment efficiently, maximizing instructional time. **Efficient assessments** minimize time away from instruction. If the teacher is busy writing long notes, he or she is not interacting with children and actively supporting their learning. In addition, if the teacher finds data collection too burdensome, he or she may not collect data often enough to assess progress. Checklists or photos usually are the most efficient methods for documenting. In order to be *efficient*, the teacher’s notes and observations should show intentionality. If the teacher is just taking copious notes without a specific intention, the notes may not be used to inform his or her understanding of the child’s progress or performance.
- Examples of efficient assessments include:
 - For a letter-naming exercise, the teacher prepares an alphabet checklist with all allows her to quickly indicate which letters the child recognizes.

- The teacher prepares a template that allows her to easily fill in the child’s responses to open-ended prompts during a book reading activity.
- Assessing a *naturally occurring activity* is often inherently more efficient—for example, teachers might find it more efficient to observe whether a child can zip a coat when getting ready to go outside rather than during a small-group activity.
- In some cases, the natural context may not be the most appropriate context. For example, the teacher may find it difficult to observe an individual child, or the teacher may have difficulty determining whether a child is actually exhibiting the skill or is simply mimicking peers. In these instances, teachers may need to assess children using a *clearly defined task* such as picture naming. However, the teacher should ensure that the task minimizes time away from instruction or is embedded within an instructional task.

	1	3	5	7	ENTER RATING
NOTE: COMPLETE '2E' FOR CHILDREN WITH SPECIAL LEARNING NEEDS OR DUAL LANGUAGE LEARNERS.					
E	Assessment tasks are not valid for child (in terms of language, culture, temperament, and/or ability); teacher does not make needed accommodations or adaptations for children with special learning needs or Dual Language Learners.	For <u>at least one</u> assessment target, teacher assesses in multiple ways, including making needed accommodations or adaptations.	For <u>some</u> assessment targets, teacher assesses in multiple ways, including making needed accommodations or adaptations.	When appropriate teacher assesses target in multiple ways. Teacher documents all accommodations or adaptations used to ensure that the child understands the task.	RATING: _____ <input type="checkbox"/> CANNOT RATE

- *Teacher assesses in multiple ways, including making needed accommodations or adaptations.*
 - To ensure that the child understands the task, to make the task valid for the child, and to honor and consider children’s differences, the teacher should use (and document) accommodations, adaptations, or prompts whenever needed.
 - To make the task valid for the child, the teacher should use tasks that are *accessible* to the child. For example, the teacher could allow a dual-language learner with a Spanish home language to respond to English prompts in either English or Spanish. The teacher may also use different languages and/or gestures. For example, if the task is rhyming, the teacher could allow the child to use nonsense rhymes or rhymes in other languages to demonstrate understanding of the concept.

- To consider and honor children’s differences, the teacher should consider the following when selecting the assessment method:
 - Linguistic differences
 - Cultural differences (for example, whether the child eats with a fork or with chopsticks)
 - Differences in temperament (whether the child will respond demonstrating what she knows and can do only in small groups or individually)
 - Differences in ability (for example, whether the child needs modifications, such as special seating, to participate in an activity)
- Specific types of adaptations and accommodations are discussed in detail on page 43.
- Note that teachers may make poor adaptations that actually distract from the task. For example, if a child is having trouble understanding the concept of rhyming, a teacher could pull out physical objects—a cat, a hat, and a ball—to help a “concrete” learner understand rhyming. In this case, however, the child may simply be confused by the physical objects, trying to find the differences/similarities in their appearance rather than their names.
- *Teacher documents accommodations or adaptations used to ensure that the child understands the task.*
- The teacher must document both the accommodation/adaptation/prompt and why it was needed. For example, “needs help holding scissors” is not providing information about an accommodation. Instead, the teacher might write, “Lacks coordination to use scissors independently; I cut the paper with her using a scissor with two sets of handles—one for me and another for her.”

	1	3	5	7	ENTER RATING
<u>FOR VIDEO-BASED ASSESSMENT AND INSTRUCTION ONLY</u>					
F	Goals of the observed assessment tasks are not consistently clear even to the observer.	Goals of the observed assessment tasks are clear to observers, but the tasks may not be described to the focal child(ren).	<u>Most</u> of the observed assessment tasks are familiar to the focal child(ren) or are clearly described.	<u>All the observed assessment tasks are familiar to focal child(ren) OR teacher ensures focal child(ren) understands tasks before beginning assessments.</u>	RATING: _____

- *Observed assessment tasks are familiar to focal child(ren) OR teacher ensures focal child(ren) understands tasks before beginning assessments.*
- To be **ecologically valid**, the observation/assessment should be familiar to the child. In other words, the task should be something that a child would be expected to do in a particular environment, rather than a task that is so unusual that the child may not

understand what is being asked of him or her (if the child does not understand the task, the child may not demonstrate knowledge or skills that he or she possesses).

- The teacher should select an assessment method that makes sense to the child; in other words, the child should understand what he or she is being asked to do. If children do not understand the task, they might not be able to show their knowledge or skills. At the lowest rating level, the goal of the task is not clear to the child and is not clear even to the observer or other adults. The teacher uses directions that are complex, use too many pronouns and/or not enough active or descriptive words to clearly communicate what is expected.
- For example, the teacher is trying to assess the child’s ability to name letters in the alphabet. He or she provides the child with an alphabet card and wants the child to name the letter associated with the picture on the card (e.g., the teacher provides a card with a picture of the sun to represent the letter “s”). The child proceeds to name the item on the card (in this case, “sun”) rather than naming the appropriate letter (“s”). Instead of assessing letter knowledge, the teacher is actually assessing the child’s ability to name items. It is clear that this child does not understand the task he or she is being asked to do.
- If the task is unfamiliar to a child, the teacher can ensure the child understands the task by providing examples and by practicing. For example:
 - If the teacher wants to know whether the child can follow directions, she could start with simple directions that the child has followed consistently in the past and use that as practice for following directions related to the assessment task.
 - If the teacher wants to know whether a child can rhyme, she could begin by demonstrating a few examples of common rhymes before asking the child to select two rhyming words out of a list of words or to produce a new rhyme.

	1	3	5	7	ENTER RATING
<u>FOR VIDEO-BASED ASSESSMENT AND INSTRUCTION ONLY</u>					
G	In the observed assessment tasks, teacher typically does not devote sufficient attention to focal child(ren) while implementing observation or assessment.	In the observed assessment tasks, teacher typically is frequently interrupted when assessing the focal child(ren).	In the observed assessment tasks, teacher typically focuses on the focal child(ren) being assessed with only small interruptions.	In the observed assessment tasks, teacher typically devotes sufficient attention to focal child(ren) while implementing observation or assessment.	RATING: _____

- *In the observed assessment tasks, teacher typically devotes sufficient attention to focal child(ren) while implementing observation or assessment.*
- Devoting sufficient attention means that the teacher is focusing all of her attention on the focal child(ren) beyond general monitoring of other children in the room. For example, the teacher should not be engaged in other activities talking with other children in the

room or texting with a parent. The teacher is unable to observe the focal child(ren) if he/she is spending time on other things.

1	3	5	7	ENTER RATING	
FOR VIDEO-BASED ASSESSMENT AND INSTRUCTION ONLY					
H	When using a standard task, does not implement in a standard way and does not note changes in administration.	When using a standard task, sometimes implements in a standard way and sometimes documents the deviations from standard administration, if applicable.	When using a standard task, usually implements in a standard way. Consistently documents any deviations.	When using a standard task, implements in a standard way (with fidelity) . Documents if concerned about valid response, if applicable.	RATING: _____ <input type="checkbox"/> NOT APPLICABLE <input type="checkbox"/> CANNOT RATE

- *Implements in a standard way (with fidelity).*
- When a teacher implements a standard task in a “standard way” (or with fidelity), he or she does so in the same way across time OR notes accommodations or deviations over time.

	NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	CANNOT RATE
I Assessments typically occur in a familiar context.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	0 <input type="checkbox"/>

- A *familiar context* is an activity setting that is familiar to the child and therefore valid.

	NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	CANNOT RATE
J Child’s family or household members help collect assessment information.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	0 <input type="checkbox"/>

- The child’s family or household members participate in documentation by recording information about the child’s progress at home and sharing this documentation with the teacher.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
K	Teacher monitors child’s progress in area of individualization with at least 3 pieces of evidence (data points).	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

The key to this item is for the teacher to show evidence that he or she is using multiple data points to monitor a child’s progress in a specific area.

- For example, the teacher may want to monitor the child’s ability to identify the letter “C.” He or she collects three different pieces of evidence to monitor the child’s progress. First, the teacher documents the child’s ability to identify the letter “C” on a letter card. Next, the teacher documents the child’s ability to identify the letter “C” at the beginning of the word. Finally, the teacher documents the child’s ability to identify the letter “C” on a poster depicting all of the letters of the alphabet.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
L	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.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

- 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.
- Consider the following graphic display:



- This teacher used a line graph that compared Ben to a benchmark for 4-year-olds showing both performance and progress in fluent letter naming (that is, naming letters quickly and easily).
- As the graph shows, in November, Ben named 10 letters—one more than the expected number of letters (9). However, his progress was slower than expected (a flatter line) between November and February. By February, his performance was below expectation: he named 12 letters instead of the expected 14. The teacher gave more attention to practicing letters with Ben, and between February and May, he made expected progress, although he was still a bit below expectation: 16 letters instead of the expected 18.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
M	Teacher continues to periodically monitor child's progress in area of individualization even after first signs of improved progress.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

- Periodically means that a teacher is monitoring child’s progress at least three times once per reporting period. The teacher should be collecting documentation on a weekly or monthly basis at a minimum, even if it is not reviewed as often.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
FOR VIDEO-BASED ASSESSMENT AND INSTRUCTION ONLY					
N	Teacher documents child behavior/performance or collects work sample.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

- The teacher will not always provide documentation or work samples from the actual video-based observation. If documentation is not provided, it should be evident that the teacher is documenting or collecting work samples on the video. You can also probe during the interview about any data collected from the video activity.

Definitions and examples for documenting the information collected

3. RATINGS FOR “DOCUMENTING THE INFORMATION COLLECTED”

When teachers implement ongoing assessments, they need to document child progress objectively, accurately, efficiently, and with relevant contextual information.

Sources: Documents (portfolios, checklists, anecdotal records), observations, interview.

- *Documentation*
 - **Documentation** is the way that a teacher records or tracks information about a child’s performance or progress over time.
 - To examine **performance**, a teacher collects data about how a child performs on a specific skill or behavior at a given point in time.
 - To examine **progress**, a teacher examines how a child’s performance changes across time (at least three times within a reporting period). A teacher may examine progress in a particular area across a brief but intense time period, such as the vocabulary a child learns about a particular month’s theme or project. Or a teacher may examine progress across the year, such as the number of letters and sounds that the child is able to recognize from month to month over the course of the year.
 - Whether examining performance or progress, the teacher needs to collect evidence to consider the child’s strengths, and areas for growth. Documentation is particularly important for examining progress, because a teacher’s memory of child performance at earlier points in the year may fade over time.
 - Examples of documentation:
 - Documentation made by the teacher
 - Anecdotal records (Anecdotal records are short written observations where a teacher records information about what a child says or does. These records should be organized in some intentional way. Often, a computer system is used to help organize such records.)
 - Checklists (By checklists, we mean a prepared list in which the teacher can identify the presence or absence of multiple behaviors, knowledge, or skills.)
 - Ratings (By ratings, we mean scales that measure along a continuum [e.g., 1-5]).
 - Rubrics (By rubric, we mean scoring guides that describe several levels and multiple aspects of performance.)
 - Pictures, videos, or audio recordings that can be annotated and used as evidence for ratings and rubrics

- Documentation made by the child
 - Work samples (such as, writing and drawing samples)
- Portfolios used by the teacher to compile various forms of documentation

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
A	Documentation provides relevant information about the general assessment context, supports, and timing.....	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

- Teachers should note general assessment information about:
 - The nature and context of the task (such as group size).
 - The date and time (to support later reflection about child progress across time).
 - Any **general assessment supports** provided to all participating children (such as prompts, environmental supports, and help from adults in the classroom). For example, if the documentation included child artwork where children had cut out a tree and pasted it on a larger piece of paper, we would want to know whether the teacher or children had cut the tree pieces.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
B	Documentation includes child-specific context, including types of prompts and supports.....	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

- Child-specific prompts may be verbal or physical.
 - **Verbal prompts** can entail asking follow-up questions or using different questions for children of differing levels of knowledge or skill. Verbal prompts can also include guidance, such as directing a child about how to reposition a puzzle piece.
 - **Physical prompts** could entail demonstrating how to write a letter or providing hand-over-hand guidance to a child trying to write a letter.
- Child-specific supports can also be verbal, physical, or environmental.
 - **Verbal support** could entail limiting a child’s choices when asking a question (“Is this wet or dry?”)
 - **Environmental support** could entail color-coding materials that go together.
- It is important for teachers to document prompts and supports so that they may later observe whether the child can succeed at a task or in answering a question without them.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
C	The documentation can be understood without the need to ask the teacher questions beyond clarifying shorthand codes.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

- A teacher’s documentation should be sufficiently clear such that another person could understand the documentation without needing to ask clarifying questions beyond shorthand codes. Consider the examples in the table below:

Documentation requires clarification	Documentation does not require clarification beyond shorthand
<ul style="list-style-type: none"> • L.: “red.” • J. ran to fence. 	<ul style="list-style-type: none"> • 10/2: [S] N. turned and elbow knocked over milk. Got rag and cleaned up. • 12/9: [F/P: TT] D. stacked 6 1” cubes. Pointed at tower. “I did 6.”

- To understand the two examples on the left, an outsider would need information about the context:
 - For “L: red,” an outsider would not know whether the child called “L” said the color red, pointed to a red object, or pointed to a red book that she wanted someone to read to her.
 - For “J ran to fence,” an outsider would not know whether the child called “J” ran to the fence for fun during free choice time or ran away when asked to come inside for a nap. The outsider also would not know whether J was running on rough or smooth terrain and whether J ran without falling.
 - Also note that neither example includes the date.
- To understand the examples on the right, an outsider would only need to clarify the teacher’s shorthand.
 - In the first example, [S] is for snack.
 - In the second example, [FP:TT] stands for [Free Play: Table Toys].

- Both of these examples have the date.

		FOCAL CHILD 1				FOCAL CHILD 2			
		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
D	Documentation includes information about the child's strengths.	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>				1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>			
		1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>				1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>			

- Including information about the child's strengths helps teachers individualize instruction. Specifically, a teacher can use a child's strengths to support their learning in other areas.
 - A child's strengths can be mentioned *explicitly* in the documentation. For example "Because Juana was able to recognize her name on the helper chart, I asked her to identify J, U, N, and A in a book. She found each one."
 - Or a child's strengths may be *implicit* in the documentation. For example "Juana identified her name on the helper chart, and during small group, she identified J, U, N, and A in a book."

		FOCAL CHILD 1				FOCAL CHILD 2			
		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
E	Documentation includes errors and misconceptions.	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>				1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>			
		1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>				1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>			

- When a teacher only documents a child's successes or whether a child was correct or incorrect, she cannot determine the issues driving low performance or lack of progress. For example, "Jenna does not find her name card" only conveys that the child was unable to identify her name. However, "Jenna picks up Julia's name card" conveys that Jenna was able to identify the initial letter of her name, but not the name in its entirety.
- Documentation can include the teacher's notes on errors/misconception or a work sample that conveys an error. For example, the teacher may provide samples of name writing collected across time showing that the child always write a certain letter backwards.
- It is also sufficient for the teacher to be able to adequately describe how he or she documents errors and misconceptions during the teacher interview even if the documentation itself is not provided.

		FOCAL CHILD 1				FOCAL CHILD 2			
		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
F	Documentation is objective, describing what happened rather than making subjective comments and judgments.	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>				1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>			

- Examples of objective and subjective documentation:

Documentation with subjective comments and judgments (inferences in italics)	Objective documentation
<ul style="list-style-type: none"> • M. <i>frustrated</i> building block tower. <i>Handled it well.</i> • Anna was <i>bored</i> and began throwing toys. • Arthur had trouble with the five-piece puzzle and <i>grew frustrated.</i> • Alex and Timothy were <i>aggressive</i> with each other in block corner. 	<ul style="list-style-type: none"> • M.'s block tower fell over. Blew out loud breath and then started rebuilding. • Anna moved from toy to toy and then began throwing them. • Arthur could not fit the pieces together and threw them down. • Alex and Timothy shouted and pushed each other in the block corner.

		FOCAL CHILD 1				FOCAL CHILD 2			
		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
G	Documentation is an intentional process, with method selected when planning instruction.	1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>				1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/>			

- Teachers are likely to collect data more often if they have efficient means of documentation. Efficiency is added by a teacher’s intentional creation of documentation. For example, a teacher may plan which skills she is attempting to observe and she may present the child with an opportunity to demonstrate skills that the teacher is interested in assessing.
- For example, when a limited number of responses are expected, the teacher can develop a checklist or series of codes to quickly and easily document responses or behaviors. Teachers may also take a picture or short video to quickly document child performance

on a task. By contrast, some teachers may keep time-consuming, running records with copious notes, some of which are never actually used to track child progress.

- Efficiency helps the teacher avoid missing critical information—a teacher writing copious notes is not able to observe carefully. By planning the specific information to collect, the teacher can easily document important context for the assessment by copying and pasting from the activity plan.

		FOCAL CHILD 1				FOCAL CHILD 2			
		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
<u>DOES THE TEACHER HAVE INDIVIDUAL LEARNING PLANS?</u>									
<input type="checkbox"/> Yes <input type="checkbox"/> No → GO TO NEXT SECTION									
H	Documentation includes child's responses/performance during individual learning plan activities.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

- The teacher should clearly document child's responses or performance during individual learning plan activities. For example, the teacher may be working with Jose to write his name. The teacher could save work samples of Jose practicing writing his name. It is important that these work samples are dated, so the teacher can assess Jose's progress with writing his name over time.

Examples and definitions for organizing the information collected

4. HOLISTIC RUBRIC, RATINGS, AND CHECKLIST FOR “ORGANIZING THE INFORMATION COLLECTED”

The teacher’s organization of the documentation should facilitate interpretation and communication with families and team members, impose a minimal burden on teachers, and provide consistent, reliable data entry.

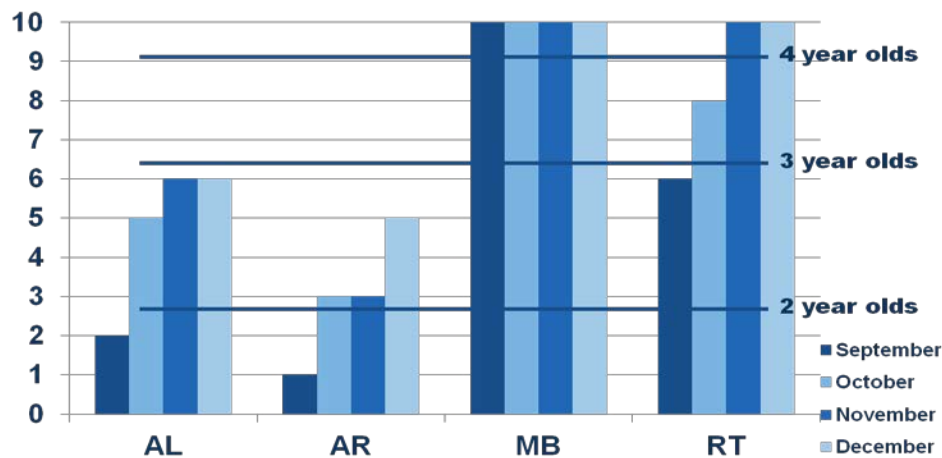
Sources: Documents (documentation from assessment, data display), teacher interview.

- *Teachers should organize their documentation to facilitate interpretation and reflection about the meaning of the data and any emerging patterns in the data. **Emerging patterns** may deal with child progress across time or with a single point in time. For example, does the time of day affect child performance? Is the child able to perform the task only in group settings, when she can mimic others? Is the child able to generalize the skill in some ways but not others?*
- *Teachers can use the organization to answer the following questions:*
 - What questions does the teacher still have about the child’s skills, knowledge, or abilities?
 - How can the teacher confirm that the child really does know what we think that he or she knows?
 - How can the teacher help the child take the next step in development?
- *The documentation should also enable the teacher to examine the progress that the child is making over time. Are there areas where the child needs additional support to make the progress expected at that age?*
- *Examples of organizational systems:*
 - Paper-based systems
 - Portfolios for compiling data from multiple sources
 - Excel spreadsheets
 - Databases like those provided by some schools
 - Graphs
 - Web-based or technology-enhanced systems to support documenting and organizing data

		YES	NO
A	The teacher views/organizes the data to compare a child's performance to a developmental expectation or benchmark for growth.....	1 <input type="checkbox"/>	0 <input type="checkbox"/>
B	The teacher views/organizes the data to compare a child's progress to a developmental expectation or benchmark for growth.....	1 <input type="checkbox"/>	0 <input type="checkbox"/>

- Teachers should organize the data to compare child progress or performance to expected growth or benchmarks. In the example below, the teacher took the rote counting information she had recorded and created a bar graph in Excel.
 - The graph allows the teacher to easily look at child progress across time and compare students.
 - In addition, the graph allows the teacher to compare a child's progress to the program expectation or benchmark for growth.
 - The three horizontal lines indicate the expectation for how far a child should be able to rote count by 2 years of age (up to 3), by 3 years of age (up to 7), and by 4 years of age (up to 10) by the start of the Head Start year.
 - In the middle of the toddler year, as is the case for these four students, the teacher would expect them to be somewhere between 3 and 7. By December, AR is halfway to meeting the 4-year-old benchmark, MB met the benchmark in September, RT met the 4-year-old benchmark in November, and so on.
- For both 4A and 4B: if the teacher's online assessment system could produce reports that view/organize the data in these ways but the teacher does not use those reports, the teacher should still get a rating of "no."

Counting to 10: Child performance in relation to benchmarks and progress over time



		YES	NO
C	The teacher views/organizes the data to compare a child's performance to that of other children in the class.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
D	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.	1 <input type="checkbox"/>	0 <input type="checkbox"/>

- In addition, the teachers should easily be able to compare a child's performance to that of other children in the class and easily identify the current level of the child's skill within a reporting period for individualization goals.
- Referring back to the "Counting to 10 graph above, the teacher is able to clearly compare how different children are progressing to the rest of the children in the class. In the graph, it is evident that RT is performing at a higher level than AR.
- The graph also allows the teacher to look at the child's own progress over time within the first reporting period of the year. In this example, the teacher is able to look at four data points per child in the fall to understand each child's individual progress.

		YES	NO
E	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).	1 <input type="checkbox"/>	0 <input type="checkbox"/>

- The teacher should consider at least 3 pieces of evidence within a reporting period to understand a child's progress on individualized goals. Examples of evidence can include:
 - Preliminary scores
 - Work samples
 - Anecdotal records

It is important to note that the pieces of evidence can be the same or a combination of the examples listed above. For example, the teacher can look at two work samples and one anecdotal record or three work samples to understand the child's progress with writing her name.

		YES	NO
F	The system for organization is efficient. (For computer systems, answer will typically be "yes" if teacher is easily able to access and enter information. Note computer application or describe below.).....	1 <input type="checkbox"/>	0 <input type="checkbox"/>

- Teachers can efficiently use electronic systems to organize information entered on a regular basis.

- If the teacher organizes the data herself, she should do so in a way that requires minimal additional time and effort. Teachers may use child portfolios with work samples collected throughout the year or summary charts. To maximize efficiency, teachers should include:
 - The date
 - Children’s names
 - Skills assessed
 - Brief notes and/or codes to indicate whether the child is having difficulty, is still in process, or has demonstrated the skill, knowledge, or behavior

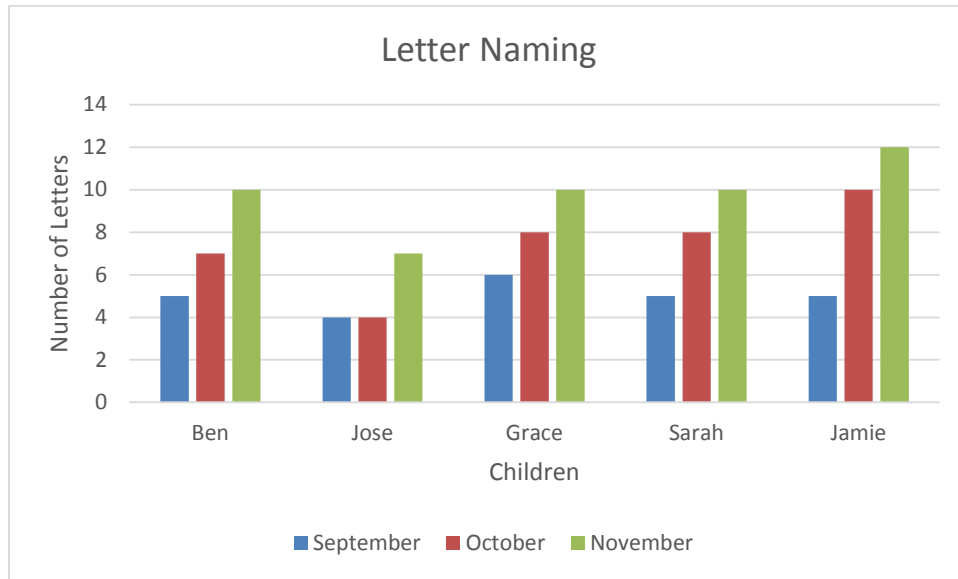
If there is no organizational system, check “no” for 4F.

		YES	NO
G	Teacher presents/organizes the data to communicate to parents about the child’s strengths and weaknesses.....	1 <input type="checkbox"/>	0 <input type="checkbox"/>

Teachers should organize the data in a way that facilitates communication with parents, so that teachers can share what they are seeing at school and parents can share what they are seeing at home. The organization should clearly identify:

- Where children are ready for more challenge or independence
- Where they might need some additional practice or other support
- Where more observations/assessments are needed to determine child performance and progress

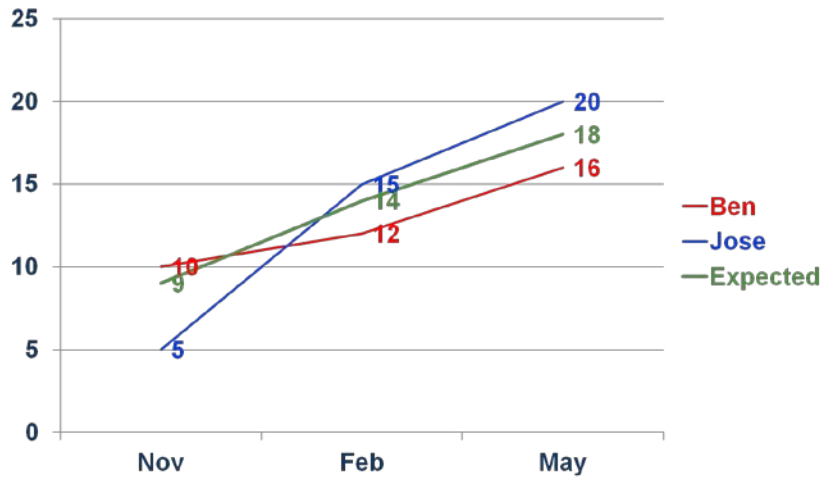
		YES	NO
H	The teacher organizes the information to look at performance by class for one or more assessment targets at a single timepoint.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
I	The teacher organizes the information to look at progress by class for one or more assessment targets across multiple timepoints.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
J	The teacher organizes the information to look at performance by subgroup for one or more assessment targets at a single timepoint.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
K	The teacher organizes the information to look at progress by subgroup for one or more assessment targets across multiple timepoints.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
L	The teacher organizes the information to look at performance by domain for the class or a subgroup for one or more assessment targets at a single timepoint.	1 <input type="checkbox"/>	0 <input type="checkbox"/>
M	The teacher organizes the information to look at a child’s progress by domain within reporting periods (multiple assessment targets across multiple timepoints).	1 <input type="checkbox"/>	0 <input type="checkbox"/>



- In the graph above, the teacher is able to look at children’s progress within one domain (in this case, the number of letters children are able to identify) at a single time point or across multiple timepoints.
- The teacher is able to look at each child’s progress both individually and in comparison to other children in his or her class. The teacher could group children by subgroup (for example, all the 3 year olds together).
- The teacher could also create a different graph that would group several skills within a single domain (for example, literacy)

		YES	NO
N	Teacher uses visual displays to depict child progress by date for at least one target.	1 <input type="checkbox"/>	0 <input type="checkbox"/>

- **Graphic displays** can help teachers assess child progress over time. Consider the example below:



- This teacher used a line graph that compared Ben and Jose to a benchmark for 4-year-olds showing both performance and progress in fluent letter naming (that is, naming letters quickly and easily).
- As the graph shows, in November, Ben named 10 letters—one more than the expected number of letters (9). However, his progress was slower than expected (a flatter line) between November and February. By February, his performance was below expectation: he named 12 letters instead of the expected 14. The teacher gave more attention to practicing letters with Ben, and between February and May, he made expected progress, although he was still a bit below expectation: 16 letters instead of the expected 18.
- Jose, on the other hand, had a slow start to letter naming: 5 letters instead of the expected 9 in November. He made greater than expected progress, however, and ended the year above expectation, with 20 letters instead of the expected 18.

		YES	NO
O	Teacher indicates in documentation when a new instructional strategy or individual learning plan is implemented.	1 <input type="checkbox"/>	0 <input type="checkbox"/>

It is important for the teacher to indicate when she makes a change in an instructional strategy or individual learning plan.

- For example, the teacher might discuss ideas for changes with her colleagues during a staff meeting at the end of each week to inform his or her individualized lesson plans for the following week. The teacher must document the decisions made to change instruction in order to evaluate the success of the instructional strategy in the future.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
P	The teacher views/organizes data in a way that makes it easy to understand the current level of the child's skill compared to prior checkpoints/reporting periods.....	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

- In the example below, the teacher created a table (that is, an ordered checklist) to understand the current level of Jose’s skill in identifying the letters in his name compared to prior checkpoints/reporting periods.
 - Initially each week and then each month, the teacher observed Jose as he wrote his name and noted the order in which he wrote the letters. On 9/8, Jose wrote the J and O in order; this seems to imply that he understood the order, but the next entry suggests that he does not really understand the sequencing yet.
- This table also shows *change over time*.
 - You can see that Jose began with only writing the O in his name. By November, he wrote all the letters in his first name in order. He maintained the skill across a month, demonstrating it again when observed in December. A teacher could easily keep this type of information on an index card and record how Jose writes his name on selected days. She might have a card for each child and observe a few children each day.
 - Alternatively, the teacher might capture this information by numbering the order in which Jose made the letters on a work sample and then taking photos of samples of work to show, across time, both the order of the letters and how he makes the letters when writing his name.

Date	Letter Recognition for Jose			
	J	O	S	E
9/4		1		
9/8	1	2		
9/21	2	1		
9/28	2	1		
10/5	1	2		
10/12	1	2	(4)	3
11/5	1	2	3	4
12/4	1	2	3	4

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	NOT APPLICABLE
Q	If the assessment system organizes the data, the teacher uses the system with fidelity—for example, the teacher enters information within a week. <i>Make notes on the time frame.</i>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	o <input type="checkbox"/>
R	If using a teacher-developed system, the teacher files or enters data on at least a weekly basis.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	o <input type="checkbox"/>

- The teacher should use the assessment system with fidelity or should organize data in his or her teacher-developed system regularly.
 - Because different program reporting requirements vary, this item requires that raters consider the guidelines provided by the assessment system being used.
 - Often, teachers are required to collect a specific number of pieces of evidence (e.g., 3 anecdotal records per week per child). Data should be entered in a way that is true to the system and not just for the sake of meeting minimum document collection requirements.
 - For example, an anecdotal note that the entire class counts to 20 during circle time does not reflect that all children are really able to count to 20. Therefore, that anecdotal note should not be used as evidence for an individual child's skills.

Examples and definitions for interpreting the data

5. HOLISTIC AND ANALYTIC RUBRICS AND CHECKLIST FOR “INTERPRETING THE DATA”

Teachers draw on multiple sources of information in making decisions about the current skills, knowledge, or behavior of children and the progress they are making. The teachers use objective, reliable data and consider alternative explanations for children’s behavior. Teachers consider the pattern of data and set hypotheses about children’s development and instructional needs.

Source: Documents, teacher interview (with questions about the video).

- To interpret the data they collect, teachers can synthesize information from a variety of sources. Specifically, teachers should interpret the data about each child’s performance and progress relative to developmental expectations for typically developing peers, benchmarks, or curricular guidelines. Teachers interpret the data in light of other available information to identify each child’s strengths and weaknesses and ultimately inform decisions about how to best support each child’s progress.

	1	3	5	7	ENTER RATING
A	Bases ratings on impressions or memory. No evidence to support interpretation.	Bases most ratings on single piece of information (that is, taken at a point in time or single event). Evidence is too scanty to support interpretation. OR Some evidence supports interpretation whereas other evidence does not.	Bases most ratings/ inferences on at least two pieces of information (can be from a single context or task). Evidence supports interpretation for most assessment targets, but evidence for some interpretations is based on inference rather than objective data. Considers alternative explanations of observed skill/behavior, including the amount of support available to the child. Considers child’s performance and progress relative to typical development or progress of peers. Examines more than one data point in between checkpoints for at least one target per child (such as preliminary ratings).	Bases most ratings/ inferences on multiple pieces of information collected in more than one context or task. Evidence is objective and supports interpretation. Considers context of observed skill/behavior. Considers child’s strengths and interests as well as challenges. Tests some alternative explanations of observed skill/behavior. Considers child’s performance and progress relative to typical development and progress of peers.	RATING: _____

- *Bases most ratings/inferences on multiple pieces of information collected in more than one context or task.*
 - The teacher should look for emerging patterns in the data collected about each child.
 - Children may perform differently on a given skill, knowledge, or behavior in different contexts or on different tasks.
 - For example, a child may be able to identify his name on his name card but not on a piece of paper with different color ink or a different size font.
 - In a group context, a child who is unable to perform a task independently may appear to perform the task in a group setting by imitating the behavior of her peers.
 - Children’s performance on a task may also depend on the time of day. If a child is tired or hungry, he may not be able to attend and organize a response as well.
 - Teachers can base inferences on data collected in more than one context or task to better understand child performance.
 - For example, when determining whether a child understands the concept of 3, the teacher might look at how the child has performed on a variety of tasks, such as asking the child to:
 - Hand her 3 blocks
 - Find the tower that has 3 blocks
 - Make a tower of 3 blocks
 - Say how many are in a set when shown sets of 1, 2, or 3 things
- *Evidence is objective and supports interpretation.*
 - The teacher’s interpretation should be supported by objective evidence (see section on objective documentation on page 23).
- *Considers context of observed skill/behavior.*
 - When teachers interpret data, they should consider the **context(s)** in which the knowledge, skill, or behavior was observed or assessed. Teachers should determine whether the child’s performance is context-specific. Does the child only exhibit the skill:
 - With specific materials or activities?
 - When the child is rested?
 - When other peers are present who are also demonstrating that behavior or skill?
 - When environmental or instructional supports are made available?
 - Teachers should consider whether the child has **generalized** the knowledge, skill, or behavior to other contexts. In other words, has the child demonstrated the skill in more than one context? For example:
 - Can the child identify rhyming words in a new finger play?
 - Does the child name letters found in words and books, as well as individual letters shown on a card?

- In the example shown below, can the child identify a single letter “s” and then also identify that letter at the beginning and end of a word?



- Teachers should also consider the scheduling of data collection. Does assessing some areas of learning at different points in the day (or week) lead to different conclusions about child progress? For example, assessing knowledge of letter names right after a group review of letters may lead to different results than assessing as children arrive on a Monday morning. Timing of the assessments could therefore present different patterns of change than would be evident if the data were collected consistently within the same context. Teachers may begin with a consistent time period and then, when children are successful in that context, examine whether they can be equally successful in other contexts.
- *Considers child’s strengths and interests as well as challenges.*
- Teachers should consider a ***child’s strengths and interests, as well as challenges***, when interpreting her observations and assessments.
 - Strengths
 - Examining what a child does well and easily may reveal skill patterns.
 - Interests
 - Interests may motivate a child to persist when faced with a challenge. What types of activities does the child select? For example, will he only sort things by color, never by shape? What does the child tend to talk about? Does the child gravitate toward activities that involve animals?
 - Interests may also indicate what the child finds challenging. A child may avoid activities that he finds difficult, indicating that the teacher may need to provide additional support or scaffolding in those areas.
 - For example, some children do not want to go to the art area because it is hard for them to cut with scissors or hold a paintbrush for a long time. They may need activities to develop strength in their hands or to orient their arm to make it easier to draw and cut. The teacher might consider whether there is other evidence that these children have problems with using their hands (for example, in eating with forks and spoons, zipping, buttoning).
 - Challenges
 - Identifying a child’s challenges across different assessment targets can help a teacher determine how to best support the child’s learning.

- *Tests some alternative explanations of observed skill/behavior.*
 - The teacher should consider (and perhaps test) **alternative explanations** for a child’s behavior, independent of whether the development was positive (that is, the child “got it right”) or the progress appears problematic.
 - For at least some dimensions, the interpretation should include setting **hypotheses to be tested in the next instructional opportunity**—for example, “Can she do it if I provide more visual cues?” or “Will he be able to do this in a different context?”
- *Considers child’s performance and progress relative to typical development and progress of peers.*
 - Typical developmental benchmarks may come from national norms for assessments.
 - Teachers should consider how the child compares to peers nationally (that is, the expectations for children of this age nationwide) and locally (that is, other children in the class).

	1	3	5	7	ENTER RATING
B	Infrequent examination of data (does not examine child’s records and make ratings at reporting periods).	Examines current data in relation to performance criteria, but does not consider the child’s rate of progress.	Evaluates progress (that is, change in performance within reporting periods). Uses data to identify areas of learning where progress is slower than expected and where it is faster than expected.	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.	RATING: _____

- *Identifies when current rate of progress has accelerated beyond expectation even if current performance is still below age level.*
 - For example, the teacher may work with children on letter naming. In December, most of the children in the class can name 12 letters except for Jose. Up until last month, Jose was only able to name one letter. Within the last three weeks, Jose learned three more letters. This information allows the teacher to look at Jose’s *rate* of progress relative to the other children. The teacher now recognizes that her strategy is working. Sufficient documentation could include a graphic display or a simple count of letter naming knowledge for each child.
- *Identifies when current rate of progress has slowed or stopped even if current performance is above age level.*

- For example, the teacher notes that Grace is already able to count to 20 while the other children are only able to count to five at the beginning of the school year. By October, most children are able to count to 12, but Grace is still only able to count to 20. Grace is not showing any progress even though she is still performing above the other children in her class.

		NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)
C	Teacher involves the family in interpreting and understanding the data.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
D	Teacher involves the other teachers/staff in interpreting and understanding the data.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

- *The teacher should involve other stakeholders, such as parents and/or other teachers, in the interpretation—for example, considering how a challenge manifests itself at home and discussing conditions under which a child does or does not exhibit a behavior or skill.*
- This must actually involve dialogue with a parent and/or another teacher about progress and performance beyond simply sharing data with a parent or colleague. The teacher should enlist the parent and/or another teacher in the process of the teacher making sense of data.

Examples and definitions for formulating instructional decisions

6. HOLISTIC RUBRIC AND CHECKLIST FOR “FORMULATING INSTRUCTIONAL DECISIONS”

Based on the interpretation of the data and any hypothesis set, teachers plan instructional activities and strategies to support increased or continued growth. They plan instructional groups intentionally to support children’s development (e.g., partnering children to provide a good language model for a child who needs to talk more, targeted instruction of small groups of children with a similar need). Instruction considers what data suggest about children’s current developmental status and progress by, for example, providing more independent practice in areas where a child is successful, testing hypotheses about the types of support children need, or considering whether children generalize skills across tasks.

Sources: Documents (data display, assessment records, lesson plans), teacher interview (with questions about the video).

Based on the interpretation of the data from observations and any hypotheses about the child’s skills, knowledge, and development, teachers plan instructional activities and strategies to support increased or continued growth. While documents collected may be considered here, raters will typically rely on interview data to rate items in this section.

	1	3	5	7	ENTER RATING
A	<p>Instruction may be designed to be interesting and engaging, but it does not draw on what is learned from the data or follow a logical plan for developing child(ren)'s skills.</p>	<p>Instructional decisions follow curriculum recommendations.</p> <p><i>Instruction includes more than one small group or individual lesson/planned interaction.</i></p> <p>Instructional grouping is based mainly on social and behavioral indicators (such as friendships or avoiding problem pairs).</p>	<p>Instructional groups and activities are designed intentionally using data to support child(ren)'s development.</p> <p>Instruction follows a logical sequence to increase child(ren)'s skills across time.</p> <p>Instructional decisions consider the evidence of current developmental status for each individual child.</p> <p><i>Some instructional decisions draw on evidence-based or professionally recommended strategies.</i></p> <p>Instructional decisions provide more intensive instruction in areas where child(ren) lag behind developmental expectations for growth.</p>	<p>Instructional decisions include modifications and adaptations based on the evidence of child development including patterns of progress across time and consideration of alternative explanations for child performance.</p> <p><i>Instructional decisions consider child(ren)'s interests and strengths as well as areas for growth.</i></p> <p>Considers child's progress relative to typical development and progress of peers.</p> <p><i>Reaches out to external resources as needed.</i></p> <p><i>Most or all instructional decisions draw on evidence-based or professionally recommended strategies.</i></p>	<p>RATING: _____</p>

- *Instruction includes more than one small group or individual lesson/planned interaction.*
- A **planned interaction** is a planned opportunity where the teacher provides a child with individual practice or specific supports; for example, planned interactions can occur during free choice time or during center time.

- *Instructional decisions consider child(ren)'s interests and strengths, as well as areas for growth.*
- Instruction should build on the child's strengths. Teachers should avoid slowing progress in areas where the child has strengths. Even when supporting areas that are challenging, teachers should still offer opportunities for continued growth in areas of strength and interest.
- *Reaches out to external resources as needed.*
- Examples of external resources:
 - Teachers
 - Coaches
 - Consultants
 - Family members of the child
- To make instructional decisions, the teacher may consult others about the conditions and contexts in which a child does or does not exhibit a specific knowledge, behavior, or skill, such as how the child performed in her previous classroom. The teacher may consult family members about whether and how a challenge in the classroom manifests itself at home. The teacher also can learn how others are supporting the child's learning.
- *Most or all instructional decisions draw on evidence-based or professionally recommended strategies.*
- Examples of evidence-based or professionally recommended strategies include:
 - Research-based sources
 - Professional organizations (such as, NAEYC or the National Council for Teachers of Mathematics)
 - Head Start and the Office of Childcare
 - A certified professional (such as a speech pathologist or behavior interventionist)
- Poor examples include:
 - Social media (e.g., Pinterest pages not associated with a professional organization)
 - Blogs

SELECT ALL THAT APPLY		DOES NOT OCCUR	OCCURS AND NOT DOCUMENTED	OCCURS AND DOCUMENTED	CANNOT RATE
B	Prompting or questioning strategies	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
C	Additional practice	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
D	Grouping strategies (homogeneity for a specific learning need)	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
E	Peer supports (heterogeneity)	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
F	Adult supports	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
G	Environmental supports (e.g., physical positioning of child)	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
H	Other adaptations or modifications designed to meet specific child needs (<i>Specify</i>): _____	0 <input type="checkbox"/>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>

- *Is there evidence of varied approaches to individualization that are responsive to the data?*
 - The teacher should **individualize in response to the data**. In other words, the teacher should ask, “What does the evidence suggest is the best approach to helping the child make the next step?”
 - The teacher should provide more intensive instruction in areas where target children lag behind developmental expectations for growth.
 - The teacher’s should use an approach repeatedly rather than just once. Specifically, the teacher should try an approach, decide whether it worked, and then move forward accordingly.
 - The teacher should collect data that clearly demonstrates whether the approach is working.
 - If you see a teacher use a strategy that is not responsive to the data, then do not include this in your rating. For example, you see evidence in the data that the child is already successful at a particular task, yet the teacher continues to prompt for support. The teacher should instead wait to allow the child to succeed independently without prompts.

The following adaptations and modifications change the task to make it easier for the child to succeed:

- *Prompting or questioning strategies*
 - **Prompts** may be used to make the task easier for a child who is experiencing challenges. Examples of prompts:
 - Breaking a task into parts and providing step-by-step instruction
 - Asking a question to help a child figure out an answer
 - Telling the child the initial sound in a word or a rhyming clues
 - Telling the child where to place a puzzle piece
 - Providing visual cues, such as helping a child notice that the words that rhyme look the same except for the very first letter
 - Providing verbal cues, such as saying “around the tree, around the tree, that’s the way to make a three” when showing a child how to draw the numeral 3 and then again when prompting them to recognize a 3 or to remember how to make the numeral 3.
 - **Questioning strategies** for a child who is strong in an area could include asking how and why questions such as:
 - Why do we need boots today?
 - What is happening outside?
 - What would happen if we went outside into the wet puddles without our boots?
- *Additional practice*
 - The teacher should increase or vary opportunities for learning and practice for children with weaknesses identified in the data.
 - A teacher can increase practice by:
 - Increasing the amount of time for practice at a single time point
 - In a small group experience, the teacher could extend the amount of time that children work on a number game or the amount of time they can practice a role in dramatic play.
 - Offering more opportunities to practice
 - The teacher could have a child practice counting during different activities throughout the day, such as counting how many children are waiting to wash their hands, how many fingers we have on one hand, the number of spoons on the table, the number of books they will read, or the number of blocks in a tower.
 - Increasing opportunities for learning by increasing the amount and length (duration) of instruction, amount of support, or frequency of instruction

- *Grouping strategies (focusing on homogeneity of skills)*
 - A teacher can strategically group children with similar learning needs to receive targeted instruction with additional practice time.
- *Peer supports (focusing on heterogeneity of skills)*
 - Examples of peer supports:
 - Planning to partner a child who is struggling with one who is performing well. For example, the teacher could partner a child who needs to talk more with one who can model good expressive language.
 - Having another child go first in an activity to show the child how to do the task.
- *Adult supports*
 - Adults can provide a child with direct assistance. For example, a teacher can use hand-over-hand assistance to help a child learn to copy a letter.

Note: Tracing a letter hand-over-hand is an adult support; if the teacher just provides tracing paper that is an environmental support (see below).
- *Environmental supports*
 - The teacher should use the evidence of the child's development (including patterns of progress across time and consideration of alternative explanations for child performance) to determine whether the child would benefit from environmental supports. The teacher can observe when a child is and is not successful and test hypotheses (such as trying different seating options) to inform the selection of environmental supports and other accommodations that may assist the child.
 - Examples of environmental supports:
 - Physically positioning a child so that he is not facing the rest of the class or so that he is sitting closer to the teacher to enable him to pay attention to instruction
 - Giving a child a stuffed animal and asking her to read to the stuffed animal
 - Using larger print or pictures with strong contrast
 - Positioning materials so that the child can reach them or see them more easily
- *Other adaptations or modifications designed to meet specific child needs*
 - Examples of other adaptations or modifications:
 - Adapting supports for children with visual impairments.
 - For example, a teacher may put each child's name on a different shape and color (maybe with the child's picture) to help children find their cubby or nametag; for a child who is blind, the teacher may use a different texture on the cubby or nametag.
 - Simplifying a task, such as
 - Having a child initially complete only the last step of a task independently and then gradually having her complete additional early steps independently
 - Teaching pre-requisite or related skills

	1	3	5	7	ENTER RATING
I	No evidence in the data to suggest a need for the individualization.	Some evidence suggesting a need, but no evidence suggesting that it is supporting the child's progress and teacher did not try different strategies (allow a 2 week trial period to look for change the strategy before trying something different). OR Some evidence suggesting a need but no individualization implemented.	Evidence that the child needed the support and is either making progress or the teacher is continuing to try different strategies.	Evidence that the teacher is using the data to make decisions about the success of strategies and changes those that are not effective in supporting the child.	RATING: _____

- *The teacher is using the data to make decisions about the success of strategies and changes those that are not effective in supporting the child.*
- The teacher may try a new strategy, appropriately document that there is no change in the child’s progress based on the data collected, and then proceed to try a new strategy.

Examples and definitions for applying instructional decisions and individualizing

7. HOLISTIC RUBRIC FOR “APPLYING INSTRUCTIONAL DECISIONS AND INDIVIDUALIZING”

The approach is responsive to the data (and progress) and the interpretation of the data. The teacher collects further evidence to evaluate whether the instructional approach is valid for meeting the targeted instructional need for this child (or these children). That is, the teacher notes instructional changes and assesses progress to evaluate the success of those changes. The teacher may use flexible evidence-based (or professionally recommended) strategies, such as using purposeful small groups, offering different questions or different levels of prompts, offering more practice for a child in a particular area throughout the day, adapting and modifying the activity, and using different instructional approaches (such as varied levels of visual or auditory cues or individual-versus-group interaction). The teacher may provide peer, adult, and environmental supports. Whenever possible, the teacher incorporates and builds on children’s strengths and interests when individualizing.

Sources: Documents (lesson plans, plans for individualized instruction), observations, interview.

1	3	5	7
<p>Provides standard instruction for every child—that is, does not make any changes for any child (no evidence of individualization or differentiation).</p>	<p>Increases opportunities for practice for children with weaknesses identified in the data.</p> <p><i>Shows evidence of an intentional adaptation for at least one child.</i></p> <p><i>Uses as least one evidence-based or professionally recommended practice.</i></p> <p>Collects at least one piece of data about a child’s response to individualized goals and instruction.</p>	<p>Increases or varies opportunities for practice for some children with weaknesses identified in the data.</p> <p>Uses some varied instructional strategies that are evidence-based or professionally recommended.</p> <p><i>Provides support for emerging skills identified in the data, recognizing when children need a challenge.</i></p> <p>Changes instruction if child is not demonstrating improvement or greater progress.</p> <p><i>Classroom instructional team has a shared knowledge about goals and instructional strategies for each child.</i></p> <p>Teacher plans and collects at least two pieces of evidence about how each child responds to individualized goals and instruction.</p>	<p><i>Increases or varies opportunities for practice or level of scaffolding for most children with weaknesses identified in the data.</i></p> <p><i>Uses varied instructional strategies that are evidence-based or professionally recommended to build on strengths and mitigate weaknesses.</i></p> <p><i>Incorporates child interests and experiences.</i></p> <p>Plans and collects at least three pieces of evidence about how each child responds to individualized goals and instruction.</p> <p><i>Organizes and reflects on 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).</i></p>

- *Shows evidence of an intentional adaptation for at least one child.*
- An **adaptation** can be increasing challenges for a child who is strong in an area and more opportunities for practice or step-by-step instruction for children who are weak in an area.
- *Provides support for emerging skills identified in the data, recognizing when children need a challenge.*
- A teacher uses peer supports to teach letter naming, grouping children in three instructional groups of mixed ability so that children with strengths in that area can assist those facing challenges. The teacher uses a weekly checklist to track which letter sounds each child can make. The teacher finds that one child in each of the groups is able to make the sounds for most or all letters in the alphabet while their peers cannot. The teacher decides to provide additional instruction for those three children by grouping them together for a weekly session on how to sound out two-letter words.
- *Classroom instructional team has a shared knowledge about goals and instructional strategies for each child.*
- Assistant teachers may assist lead teachers in collecting, documenting, organizing, and interpreting ongoing assessment data. Consequently, both the lead teacher and any assistant teachers should be aware of each child's goals and the instructional strategies selected to help the child achieve those goals.
- *Increases or varies opportunities for practice or level of scaffolding for most children with weaknesses identified in the data.*
- Examples of opportunities for practice:
 - **Additional practice** on a task with adult support, either individually or in small groups
 - **Peer strategies** to offer practice and reinforce skills, such as pairing a child who is struggling with a more competent and supportive peer
 - For example, in a matching game where the children take turns matching objects that rhyme or that are used for the same function (car to a bus; a saw to a pair of scissors), the teacher might encourage the more verbal child to take the lead in talking about why the objects go together while the less verbal child continues to help make the matches.
 - **Embedded opportunities** for the child to briefly practice a skill repeatedly throughout a day. Examples of embedded opportunities for practice:
 - Asking a child to count with one-to-one correspondence in a game in a small group
 - Asking a child to count:
 - Napkins for their table at snack time
 - Steps to go to the circle area
 - The number of boys and the number of girls in their line at the sink while waiting a turn to brush teeth
 - The number of blue toothbrushes and red toothbrushes

- Reading a counting book, such as *The Doorbell Rang* in the story corner
- *Uses varied instructional strategies that are evidence-based or professionally recommended to build on strengths and mitigate weaknesses.*
- Examples of evidence-based or professionally recommended strategies:
 - Using flexible small groups to intentionally group children in ways that address various learning needs and strengths
 - Individualizing instruction within a group by offering different questions or levels of prompts to each child
 - Individualizing instruction by offering more opportunities for a child to practice a particular skill throughout the day
 - Adapting and modifying an activity
 - Using different instructional approaches (such as varied levels of visual or auditory cues or individual-versus-group interaction)
 - Providing peer, adult, and environmental supports
- *Incorporates child interests and experiences.*
- As mentioned above, a teacher can use a ***child's interests*** to motivate them to persist in the face of difficulty.
- Teachers can also build on a ***child's experiences***, including their strengths, to help them succeed in the future.
 - For example, if a child gravitates toward certain approaches to succeed at tasks, the teacher can use those approaches to learning to help them learn in other areas. For example, does the child learn better with visual cues? Does he retell only the parts of the story that are shown in the pictures? Or does he respond better to sound cues, like providing the initial sound of the word to prompt for an answer? What about tactile cues, like tracing a shape to learn how to make it? Does the child pay attention to color, shape, sound, touch or movement?
 - Likewise, a child's strengths can also help the teacher identify which processes are easier for the child. For example, is it easier for the child when she sees the big picture, or does she focus on the parts of things? Does she remember information from songs or information that she talks about or repeats? How much time does she spend practicing the skills that she has now mastered?
- *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).*
- Teachers should collect data often enough to determine whether the instructional strategy is influencing the child's progress within a reporting period. If growth is flat or negative, the teacher should change the approach.
- For example, consider a child who is struggling with letter writing. The child's teacher has been collecting weekly data on the number of letters the child is able to write and finds that the child's rate of growth has been lower than expected. The teacher decides to offer the child an environmental support – tracing paper with letters written in yellow

highlight. The teacher collects data on the child's progress twice weekly for three weeks and finds that the child's rate of growth in letter writing is not improving. She then switches to providing an adult support – hand-over-hand tracing. The teacher again tracks the child's progress twice weekly over three weeks and finds that the child's rate of growth has increased.

Examples and definitions for characteristics ratings

Items 8A-8P contain a series of items which are rated on a 1-4 scale with 1 being almost never occurring and 4 being almost always evident (see general guidelines below). The assessor will determine how characteristic each item was based on his or her review of the documentation, video observations, and teacher interview. These final items make use of the same definitions previously described in this appendix.

GENERAL GUIDELINES FOR RATING CHARACTERISTICS		
Rate how characteristic the statement is of this teacher/classroom.		
1	Not At All (almost never)	You did not see this at all, or it was seen only once or twice and it is a behavior/type of documentation that usually occurs frequently in classrooms.
2	Minimally characteristic (sometimes evident)	Rate here if you see it occasionally and it is appropriate at many other times, or you see it happen sometimes, and it is only partially present (for example, some information about context on a few pieces of documentation but never complete documentation) or present only for a single assessment target/learning objective.
3	Strongly characteristic (frequently evident)	Rate as 3 if something happens frequently and across domains (as appropriate), but does not occur at all the appropriate times. Also rate here for something that would normally be 'low frequency' but is evident at many of the appropriate times or in most of the appropriate documents.
4	Extremely characteristic (almost always evident)	Something that happens at appropriate times and in appropriate documents across multiple domains. If something happens frequently but is missing in some appropriate documents or observations, then code as 'strongly characteristic' rather than 'extremely characteristic'.

This page has been left blank for double-sided copying

APPENDIX F

SAMPLE TRAINING AGENDA

This page has been left blank for double-sided copying

SAMPLE TRAINING AGENDA

- Independent advance reading materials (3 hours)
 - Justification for the measure and conceptual framework (Chapter I of final report)
 - Description of the EDIT measure and its administration (Chapter II of final report)
 - EDIT instrument (Chapter II of final report)
 - EDIT interview protocol (Appendix A of final report)Head Start Early Learning Outcomes Framework (<https://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/approach/pdf/ohs-framework.pdf>)
 - Manuals or other materials for the ongoing assessment system used by teacher(s) or program(s) of interest (for example, Teaching Strategies GOLD)
- Training session 1 (3 hours)
 - Introduction to the EDIT
 - Conceptual model
 - Purpose of the EDIT
 - Overview of the three methods
 - Summary of the measure
 - Administration procedures
 - Overarching definitions from training manual
 - Training manual definitions and examples (Appendix E) for:
 - Selecting the assessment target section
 - Selecting and implementing the assessment method section
 - Documenting the information collected section
- Training session 2 (3 hours)
 - Training manual definitions and examples for:
 - Organizing the data section
 - Interpreting the data section

-
- Formulating instructional decisions section
 - Applying instructional decisions and individualizing section
 - Characteristics ratings section
 - Independent training exercise (3 hours)
 - Trainees review sample documents, video recordings, and a mock teacher interview for a single teacher and rate using the EDIT measure.
 - Training session 3 (3 hours)
 - Trainees and trainers discuss training exercise by going through each item in the EDIT
 - Trainees present rationale behind ratings
 - Trainers provide corrective feedback and explanation as needed
 - Trainees practice administering portions of the interview
 - Certification
 - Training would typically conclude with certification according to a certification plan that specifies the necessary threshold of agreement between the trainee and trainer EDIT scores. Certification could be done during in-vivo visits pairing a trainee with a trainer or using materials created for certification and scored by a trainer(s).
 - In addition, trainers may want to consider evaluating trainees on their ability to conduct the teacher interview.

APPENDIX G

ADDITIONAL METHODS CONSIDERED

This page has been left blank for double-sided copying

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 study 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 study 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.

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

