Tailored Teaching: Teachers' Use of Ongoing Child Assessment to Individualize Instruction

Volume I: Conceptual Framework and Measurement Plan

OPRE Report 2014-14
June 2014

Tailored Teaching: Teachers' Use of Ongoing Child Assessment to Individualize Instruction

Volume I: Conceptual Framework and Measurement Plan

OPRE Report 2014-14

June 2014

Sally Atkins-Burnett, Shannon Monahan, Lauren Akers, Judith Carta, Barbara A. Wasik, Kimberly Boller

Contract Number: HHSP23320095642WC/HHSP23337041T

Mathematica Reference Number: 40158

Submitted to:
Jennifer L. Brooks, Kelly R. Fisher, and Nina Phillipsen 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, SW
Washington, DC 20447

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

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

Suggested citation:

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, and Evaluation, 2014.

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 report and other reports sponsored by the Office of Planning, Research and Evaluation are available at http://www.acf.hhs.gov/programs/opre/index.html.

MATHEMATICA Policy Research

TEMPLE UNIVERSITY College of Education

KU

OPRE

ACKNOWLEDGMENTS

This report would not have been possible without the assistance of many people. First, we wish to thank Jennifer Brooks, our federal project officer at the Office of Planning, Research, and Evaluation (OPRE), as well as OPRE Society for Research in Child Development Fellows Nina Philipsen Hetzner and Kelly Fisher, for their oversight, guidance, and vision. We are also grateful for the ongoing input of Amanda Bryans from the Office of Head Start (OHS). Other staff members from OHS and OPRE provided us with valuable feedback about the needs of the national Head Start program and the research community. In addition, we were fortunate to have the expertise of our expert panel members, whose guidance on the topic of ongoing assessment and individualization was vital.

Finally, we gratefully acknowledge our colleagues at Mathematica Policy Research who contributed their time, effort, and expertise to this project. These staff include Felicia Hurwitz, who served ably as project manager; Claire Smither Wulsin, who was integral to the literature review; Emily Moiduddin and Nikki Aikens, who provided thoughtful quality assurance; and Kelly Kramer and Amanda Bernhardt, who provided editorial support. Barbara Collette and Karen Groesbeck formatted and produced both volumes of this report.

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

This page is left intentionally blank

EXPERT PANEL MEMBERS

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

Stephen Bagnato
University of Pittsburgh

Linda Broyles
Southeast Kansas Community Action Program

Virginia Buysse
FPG Child Development Institute

Lynn Fuchs
Vanderbilt University

Leslie Nabors Oláh
Educational Testing Service

Sheila Smith
Columbia University

Patricia Snyder
University of Florida

This page is left intentionally blank

CONTENTS

OVERVIEW xiii

EXECUTIVE SUMMARY xv

I INTRODUCTION AND ROADMAP TO THE REPORT 1

A. Building a Shared Understanding: Ongoing Assessment and Individualization 1

B. The Importance of Developing a Measure of Early Childhood Teachers' Use of Ongoing Assessment to Individualize Classroom Instruction 3

C. Literature Review 4

D. Roadmap to the Report 5

II CONCEPTUAL FRAMEWORK 7

A. Assumptions Underlying the Conceptual Frameworks 7

B. A Theory of Change 8

C. The GOM Approach 10

D. Conceptual Framework for the GOM Approach 12

E. The Curriculum-Embedded Approach 16

F. Conceptual Framework for the Curriculum-Embedded Approach 18

G. A Focus on the Curriculum-Embedded Approach 26

III PROPOSED MULTI-METHOD MEASURE 27

A. The Need for a Multi-Method Approach 27

B. Measurement Issues 31

1. Balancing Validity, Reliability, and Feasibility 31
2. Accommodating a Variety of Assessment Systems 33
3. Issues Related to Selection of Classrooms and Scheduling 33

C. Recommended Multi-Method Approach 34

1. Document Review 35
2. Video-Based Classroom Observations 39
3. Teacher Interview 40

D. Initially Targeting Two Domains 43

E. Summary of the Recommended Multi-Method Measure 44

F. Optional Approaches to Inform Professional Development 45

1. Pedagogical Content Knowledge (PCK) with Scenario Probes 45
2. Standard Pedagogical Task 45
3. Advantages and Challenges of Optional Approaches 46

G. Complementing Other Classroom Observation Measures 46

IV CONCLUSIONS AND NEXT STEPS 47

APPENDIX A: DEFINITIONS OF KEY DATA SOURCES AND METHODS

APPENDIX B: EXAMPLES FROM THE LITERATURE OF PRIOR USE OF DATA SOURCES

APPENDIX C: STEPS FOR PRE-TEST

APPENDIX D: TEACHER QUESTIONNAIRE: ORAL LANGUAGE & COMPREHENSION

TABLES

IV.1. Rounds of Data Collection for the Pre-test 49

EXHIBITS

I.1. Glossary of Key Terms as Used in the T3 Measurement Plan 2

I.2. Key Findings from the Literature Review 5

III.1. Examples of Contextual and Background Information to Gather in a Teacher Interview 30

III.2. Examples of Documentation 35

III.3. Examples of Questions to Be Addressed by the Document Coding Scheme, by Stage of the Conceptual Framework 37

III.4. Rubric to Examine the Selection of Assessment Targets 38

III.5. Checklist to Examine Lesson Plans for Evidence of Applying Instructional Decisions and Individualizing 39

III.6. Rubric to Examine Teacher Interview Responses and Lesson Plans for Application of Instructional Decisions and Individualization 43

III.7. Example of a PCK Assessment Item 45

FIGURES

ES.1. Conceptual Framework for Curriculum-Embedded Approaches xvii

II.1. Assessing Teachers' Use of Ongoing Child Assessment to Individualize Instruction: Theory of Change 9

II.2. Conceptual Framework for the GOM Approach 13

II.3. Conceptual Framework for Curriculum-Embedded Approaches 19

III.1. Multi-Method Measure Model: Document Review 35

III.2. Multi-Method Measure Model: Video Observation 39

III.3. Multi-Method Measure Model: Teacher Interview 40

IV.1. Two Alternative Methods for the Focus and Frequency of Pre-test Video Recordings 50

This page is left intentionally blank

OVERVIEW

In 2012, the Office of Planning, Research, and Evaluation within the Administration for Children and Families, U.S. Department of Health and Human Services, 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 is twofold: (1) to develop a conceptual framework of early childhood teachers' use of ongoing child assessment to individualize instruction and (2) to create a measure to examine this process. Besides defining the key aspects of ongoing assessment and individualization, we are also seeking ways to efficiently examine how they are carried out. The ultimate goal is to help teachers conduct ongoing assessments, use the assessment data to "individualize" instruction for each child, and enhance their children's outcomes.

We began by conducting a literature review (see Volume II of this report), using our findings to develop conceptual frameworks for ongoing assessment and individualization. These models focused on the two most common approaches to ongoing assessment used in early childhood classrooms: general outcomes measures and curriculum-embedded approaches. In each model, we identified several indicators of quality to consider when measuring how well a teacher used ongoing assessment to individualize instruction; these indicators served as the foundation for the development of a measurement tool called the Tool for Tailored Teaching (T3). This tool, once fully developed, will analyze how teachers use curriculum-embedded approaches to tailor instruction for specific children.

The T3 will be a multi-method measure—that is, it will use several methods to gather data: a document review, video-based observations, and a one-hour teacher interview. Teachers will video-record their assessments over a two- to three-week period, followed by a one-day visit from researchers to conduct the document review, rate the videos, and interview the teachers. We will develop scoring systems for each data source, including holistic rubrics, ratings, and checklists.

In the earliest stages of development, we will conduct an iterative pre-test to hone the T3 and assess its feasibility. The pre-test will consist of three rounds of data collection, ultimately including five centers and 10 classrooms that use ongoing assessment systems.

Our findings from this study can fill a critical knowledge gap in the early childhood field, particularly in Head Start: how do teachers use ongoing, curriculum-embedded assessment approaches to deliver high-quality, individualized instruction? The T3 will reveal how teachers use these assessment data to understand children's development and to individualize their instruction. At first, the T3 will primarily be used by researchers to help Head Start staff and early childhood professionals understand how teachers use ongoing assessment. But eventually, teachers, mentors, and coaches will be able to use it to discern the strengths and weaknesses of their programs' ongoing assessment practices. This could ultimately lead to more effective practices in the classroom and better outcomes for children.

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."[1] The purpose of the project is twofold: (1) to develop a conceptual framework of early childhood teachers' use of ongoing assessment to individualize instruction and (2) to create a measure to examine this process. This project can make a significant contribution to the early childhood field, both by defining the key aspects of ongoing assessment and individualization and by providing guidance on efficiently measuring how they are carried out. Ultimately, it may help teachers enhance their practices in the classroom which, in turn, will improve early childhood program performance.

Head Start recognizes the importance of using ongoing assessment to individualize instruction for each child. Over the past five years, the Office of Head Start (OHS) has elaborated on its vision for preschool child and family outcomes, strengthened its focus on monitoring program and classroom quality, and developed tools to support ongoing assessment in daily practice (U.S. Department of Health and Human Services 2010; Atkins-Burnett et al. 2009). Recently, the Advisory Committee on Head Start Research and Evaluation, convened by the secretary of the U.S. Department of Health and Human Services, advocated investing in evidence-based and data-informed practices across all domains of teaching and learning (Advisory Committee on Head Start Research and Evaluation 2012).

Despite the importance of using ongoing assessment data to guide instruction, and the Head Start program requirements to do so, information is sparse on how early education teachers actually collect and use these data to tailor their instruction. Policymakers, practitioners, and researchers continue to see an urgent need for research in this area in the quest for better educational outcomes (Bambrick-Santoyo 2010; Black et al. 2003, 2004; Fuchs and Fuchs 2006; Hamilton et al. 2009; Marsh et al. 2006).

A. Two Approaches to Ongoing Assessment

Ongoing assessment of children's progress is increasingly a priority in early childhood classrooms, yet teachers' use of these assessments has not been extensively researched. General outcomes measures (GOMs) and curriculum-embedded approaches are two common approaches to ongoing assessment used in these classrooms.

General outcomes measures. In the GOM approach, teachers use a brief measure with strong evidence of reliability and validity to conduct frequent, standard assessments of children's progress toward a long-term goal. Central to this approach is the repeated measurement of a few key skills that represent the entire skill set required to achieve a given goal, rather than measuring the full skill

[1] This project focuses on all forms of ongoing child assessment, of which child progress monitoring is a common form. The content of this report will be broader than the title of the project implies. For more information, see the section in Chapter 1 of this report entitled, "The Importance of Developing a Measure of Early Childhood Teachers' Use of Ongoing Assessment to Individualize Classroom Instruction."

set. A child's increasing proficiency on a GOM is indicated by improved performance on these same skills measured over time.

With GOMs, children's performance may be measured as infrequently as three times per year or as often as once per week (Jenkins et al. 2009). The probes to obtain these performance samples typically range from one to five minutes, depending on the outcome (that is, the knowledge, skill, or behavior) being measured. One common application of GOMs is Response to Intervention (RTI)—an approach to early intervention involving the regular screening of all children throughout the year. Children not progressing as expected receive intensive support as well as frequent assessments to test whether the support is helping (Hamilton et al. 2009; National Association for the Education of Young Children et al. 2012; Buysse and Peisner-Feinberg 2013). GOMs typically do not focus on the full set of child outcome domains. Most GOMs in preschool currently focus on language and literacy, and some focus on mathematics.

Curriculum-embedded approaches. The most commonly used systems for assessing the progress of children in early care and education are curriculum-embedded approaches. These assessments are used to examine children's progress relative to early learning standards and the skills and knowledge taught via a specific curriculum. Teachers using this approach often collect assessment information as they are teaching their normal curriculum. The assessment tasks are intended to be authentic in context; that is, they are "opportunities created for children that reflect typical experiences rather than discrete isolated tasks that are irrelevant to the child's daily life" (Pretti-Frontczack et al. forthcoming). Some curriculum-embedded approaches are developed by the curriculum developers to align closely with the material being taught ("curriculum-based assessments" such as the Teaching Strategies: GOLD assessment used with the Creative Curriculum), whereas other such assessments are derived from national standards and developmental expectations ("curriculum-embedded assessments" such as Galileo and the Work Sampling System).

Teachers typically assess children's performance in relation to criteria on rubrics provided by the assessment system. These rubrics specify different levels of performance based on end-of-year goals, but often provide no guidance regarding children's expected progress throughout the year. In addition, although the tasks being assessed are embedded within daily activities and aligned with curriculum goals, the tasks are not standardized and require teachers to collect assessment data from multiple sources. The assessments may use a variety of data collection methods, such as observation recording forms, worksheets, standardized assessments, and portfolios.

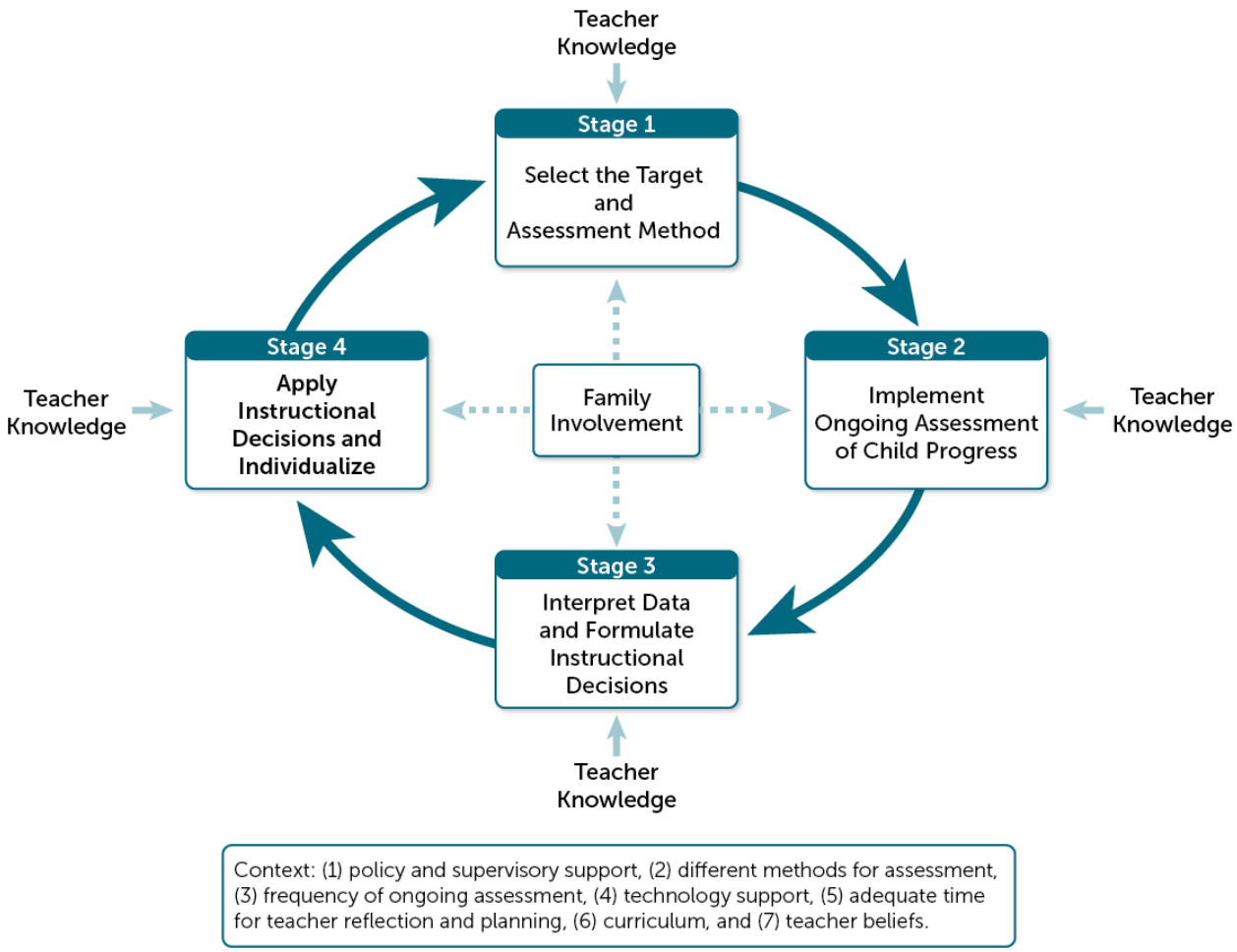
B. A Focus on Curriculum-Embedded Approaches

This report presents a measure development plan for curriculum-embedded approaches. We chose to focus on these assessments 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, as they traditionally cover several domains of development.

Figure ES.1 shows the conceptual framework for using curriculum-embedded approaches to monitor children's progress on an ongoing basis and individualize instruction. The model has four iterative stages: (1) selecting the assessment target and method; (2) implementing the assessment; (3) interpreting the assessment data, including hypothesis setting and selection of instructional

decisions; and (4) applying instructional decisions. We explain each stage and its quality indicators below.

Figure ES.1. Conceptual Framework for Curriculum-Embedded Approaches



Stage 1: Selecting the target and assessment method. The assessment system is often selected by program staff rather than by teachers. However, teachers have some autonomy in selecting the assessment target (the skill, knowledge, or behavior to be assessed) and the assessment method (how that skill, knowledge, or behavior will be assessed, such as observations, structured tasks, or standard tests), although both are also influenced by the selected assessment system. In choosing assessment targets, teachers consider the desired end-of-year outcome and set targets that track progress toward that outcome. They should collect, interpret, and reflect upon the data and make instructional adjustments throughout the year, frequently enough to monitor and guide children's progress. In this way, teachers' decisions play a large role in the assessment process throughout the entire year. The supports available to the teacher within different assessment systems vary, as does the strength of the link to the curriculum.

When examining teachers' selection of assessment targets and methods, there are several indicators of quality to consider. These include the assessment target's relation to meaningful and developmentally appropriate outcomes, key behaviors, knowledge, or skills; the ability to affect the assessment target through intervention and change; the alignment of the target with the curriculum

and method of data collection; the focus on observable and generalizable behaviors; the validity of the data collection method (including linguistic and cultural appropriateness); and the efficiency and feasibility of frequent data collection.

Stage 2: Implementing ongoing assessment. Typically, the teacher will weave assessment tasks into his or her instructional activities. Efficient assessment is therefore important to making the most of instructional time. The assessment should also be ecologically valid, using tasks that fit into a child's usual routine in class or at home. Teachers must document child progress objectively, accurately, and with relevant contextual information.

The indicators of quality to measure at this stage include the soundness of the assessment itself (whether it is ecologically valid, appropriate for the child, and fair) and the teacher's approach to documenting the findings (whether it is efficient, consistent, and objective, taking contextual factors into account).

Stage 3: Interpreting data and formulating instructional decisions. In an ongoing assessment system, teachers need to be able to interpret the data about each child's performance compared with performance expectations, which are usually based on developmental or curricular guidelines or the scores of typical same-age peers. Teachers may also need to combine the assessment data with other relevant data. The data are used to identify children's strengths, weaknesses, interests, and learning differences; based on the findings, teachers then select the best way to support each child's progress. This process may be conducted in teams with the support of other teachers, coaches, consultants, and family members.

When measuring quality at this stage, researchers should consider whether the assessment data are organized so as to (1) facilitate interpretation and easy communication with families; (2) impose a minimal burden on teachers; and (3) provide consistent, reliable data entry. Teachers' reflection on and interpretation of the data should also be evidence-based and consider alternative hypotheses.

Stage 4: Applying instructional decisions and individualizing. A requirement of the Head Start Performance Standards, individualization is important for maximizing child progress (*Federal Register* 2011). It involves planning and delivering high-quality, evidence-based instruction that is targeted to individual children and reflects the data collected about each child.

The indicators of quality to measure at this stage include the use of evidence-based strategies that are responsive to the data, implemented with fidelity, and evaluated in an ongoing manner. Teachers should also individualize their lessons using a variety of approaches, while building on children's strengths and interests.

Personnel, family, and contextual factors affecting most stages. The conceptual framework contains several factors that could potentially affect the implementation quality across stages. These factors include teachers' knowledge and beliefs about assessment, instruction, and children's development as well as families' involvement in the assessment process. In particular, each stage of the process calls for a specific kind of teacher knowledge. Stage 1 calls for teacher knowledge of assessment and child development, for example, whereas stage 3 calls for teacher knowledge of instruction, pedagogy, and child development. If a teacher has more knowledge in one area than in another, her quality of her implementation may vary across stages. The context in which the assessment occurs also affects the quality of implementation. Context includes key aspects of the program structure that help or hinder teachers' use of ongoing assessment, such as:

- The degree of policy and supervisory support for ongoing assessments
- The availability of adequate time for reflection and planning
- A culture of using data to inform instructional planning and opportunities for teachers to collaborate as they assess and interpret data
- Access to professional development opportunities and information about evidence-based or professionally recommended instructional strategies

C. Proposed Multi-Method Measure

The indicators of quality identified in the conceptual framework served as the foundation for the development of a plan to measure teachers' use of ongoing assessment to individualize instruction. This measurement plan draws on information from a number of sources, including a literature review, input from ACF and an expert consultant group, examples of teachers' assessment documentation, and reviews of manuals for ongoing assessments. Together, these sources helped us identify key constructs for measurement as well as data sources for measuring these constructs.

We propose using a multi-method measure—the Tool for Tailored Teaching (T3)—to assess teachers' use of ongoing assessment data. The T3 will draw on three data sources: a document review, video-based observations, and a one-hour teacher interview with a reflective "think-aloud" protocol.[2] Collecting data from these sources, which will provide both overlapping and distinct information, will be critical to understanding all aspects of the assessment process. Teachers will video-record certain assessment and instructional activities over a two- to three-week period, followed by a one-day visit from researchers to conduct the document review, rate the videos, and interview the teachers. We will develop scoring systems for each data source. The T3 is designed to capture the constructs in each stage of the conceptual framework while balancing the competing considerations of (1) the measure's reliability and validity; (2) the burden placed on researchers, teachers, and classrooms; and (3) budgetary limits.

Document review and ratings. For the document review, researchers will gather ongoing assessment data (such as a portfolio) for two children, one who is performing well and another who is struggling, to see how teachers are actually using the data to individualize their instruction for each child.[3] The researchers will also review the teachers' lesson plans for evidence of individualization. Rubrics, checklists, and ratings will be used to evaluate each document.

Video-based observations. The teacher will video-record a combination of assessments and small-group instruction that includes one or both of the target children. The researchers will watch the video after rating the documents and analyze it using rubrics, checklists, and ratings.

Teacher interview with reflective think-aloud protocol. During the one-hour teacher interviews, the researchers will ask for additional details on the documents and videos as well as

[2] Teachers will be asked to reflect and "think aloud" about how they made decisions as they conducted assessments and used the data to inform their teaching. Throughout this volume, we use the term "think-aloud" to refer to this reflective process.

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

teachers' planning for and use of adaptations, modifications, and individualized teaching strategies. Teachers will describe how they use the data (for example, to determine whether children are making adequate progress, to make instructional decisions, or to involve families). Researchers will rate this information using a rating or rubric, coding the teachers' responses about their interpretations of the data, any alternatives considered, their decisions about how to individualize, and the success of their efforts.

D. Pre-testing

Pre-testing will encompass the earliest stages of developing the T3 and testing its feasibility. The pre-test will consist of three rounds of data collection, with visits to five centers and 10 classrooms that use ongoing assessment systems. During this iterative process, we will try two different approaches to the focus and frequency of the video recording. After each round of data collection, the full team will debrief and consider changes to the items, protocols, and procedures. ACF and the expert consultants will have the opportunity to review the recommended changes, and we will incorporate their feedback.

For the pre-test, we propose focusing on two domains: (1) language and literacy and (2) social and emotional development. Per the expert panel's recommendation, we suggest limiting our focus to two domains to ensure the feasibility of measure development within the scope of this project. We chose these domains due to their prevalence in early childhood curricula and the likelihood that we might observe variability in the use of assessment data to tailor instruction, which would enable us to see if our measure can capture differences in the quality of teachers' implementation. Limiting the number of domains will also help narrow our focus, allowing greater opportunity to refine the measurement of one or two domains and stay within our budget. However, please note that we will also include items in the T3 that examine whether teachers are drawing on information across domains to interpret and use the assessment data.

E. Potential Uses

By exploring the role of ongoing assessment in delivering high-quality, individualized instruction, this project can significantly strengthen the knowledge base for early childhood education, especially for Head Start. The T3 in particular could provide valuable information about how teachers use curriculum-embedded assessment data to understand children's development and to plan their instruction.

The final, validated version of the T3 could be used by researchers, sponsoring agencies, administrators, teachers, mentors, education coordinators, and coaches from individual programs or by networks of programs for an array of purposes. Researchers will be the most frequent T3 users at first, as they help staff at Head Start and in the early childhood field understand how ongoing assessments is used in classrooms. The experts we spoke with repeatedly noted that we lack even basic information about whether and how early childhood teachers use ongoing assessment to individualize instruction. Consequently, the T3 offers an unprecedented opportunity to inform the early childhood field's basic understanding of a process that is valued and even mandated but previously has not been measured. The T3 could also inform teachers, mentors, and coaches about the strengths and weaknesses of their programs' use of ongoing assessment. This could lead to teaching that is better adapted for each child and, ultimately, to stronger outcomes for all children.

CHAPTER I. INTRODUCTION AND ROADMAP TO THE REPORT

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."[4] The purpose of the project is to (1) develop a conceptual framework of early childhood teachers' use of ongoing child assessment to tailor their instruction for each child and (2) create a measurement tool to examine this process. This project could make a significant contribution to the early childhood field, both by defining the key aspects of ongoing assessment for individualization and by providing guidance on efficiently measuring the use of these systems. Ultimately, it may lead to better teaching practices that will improve early childhood programs.

In this chapter, we begin by defining the terms "ongoing assessment" and "individualization" as used in this report as well as providing a glossary of other key terms. We then highlight the need for a conceptual framework and a measure of teachers' use of ongoing assessment to individualize instruction. We describe key findings from a literature review that have informed our measurement development, and we conclude by explaining the structure of this report.

A. Building a Shared Understanding: Ongoing Assessment and Individualization

Because the terms "ongoing assessment" and "individualization" are used in specific ways in different contexts, we begin by articulating our operational definitions of each. We then provide a glossary defining other key terms used in this report in Exhibit I.1.

Ongoing assessment. Ongoing child assessment refers to the "continuing observation and documentation [that] teachers complete to determine whether teaching practices need to be adapted to better meet children's needs" (National Center on Quality Teaching and Learning 2012). In other words, teachers use ongoing assessment to measure children's performance, including their skills, knowledge, and behavior; track their progress over time; and tailor instruction to meet each child's needs. The goals of this type of assessment vary. In this context, the goals are to (1) inform the teacher's instruction for the entire group as well as for each child, (2) determine whether current instructional approaches are helping children become ready for school, and (3) identify the appropriate level of additional support or type of instructional modifications needed. Overall, in early childhood settings, the information from assessments should be used by teachers to assess progress and then support children's learning.

Individualization. Individualizing means providing instruction that is adapted to each child's unique strengths and challenges. Teachers can individualize their lessons by adapting instruction or content (curriculum), increasing opportunities for the child to practice a certain skill, and providing environmental or other supports to better meet the needs of each child. In this report, "individualization" specifically refers to the process of using data to identify a child's skill or ability level in a given area and to determine and use the teaching practices needed to support the child's

[4] This project focuses on all forms of ongoing child assessment, of which child progress monitoring is a common form. The content of this report will be broader than the title of the project implies. For more information, see the section in Chapter 1 of this report entitled "The Importance of Developing a Measure of Early Childhood Teachers' Use of Ongoing Assessment to Individualize Classroom Instruction."

learning. Teachers gather data as part of a continual process to assess the child's response to instruction and make adjustments to ensure the child's growth.

Exhibit I.1. Glossary of Key Terms as Used in the T3 Measurement Plan

Term	Definition
Assessment system	A system may include the defined goals, standards, or objectives being assessed; the methods used to assess; and guidance provided for organization and interpretation of data. It may also include materials for communicating assessment results with families and others, and recommendations for next steps in instruction or when and how to intervene.
Construct*	The concept or characteristic of a child, teacher, or classroom that an assessment is supposed to measure
Fidelity of implementation	The degree to which an intervention, program, curriculum, or assessment is delivered as intended
Generalizability	The ability to produce the similar results under different conditions
Probe	An additional question to better understand the teacher's thinking, decision-making, understanding, or interview responses
Reliability*	The extent to which scores obtained from an assessment or group of assessments are accurate and consistent over one or more possible sources of error, including time, raters, items, environment, and sample groups of a population
*Internal consistency reliability**	A measure of the reliability of a score derived from the relationship among items of a single instrument and their ability to measure the same construct. Internal consistency reliability is presented as the correlation between groups of items or among all items.
*Rater agreement**	The extent to which different raters or observers obtain the same information; it can include agreement on scoring of items, administrative procedures, or observation of a given behavior.
*Test-retest reliability**	The stability of test results over time
Validity*	The degree to which an assessment accurately measures what it is designed to measure
*Concurrent validity**	Demonstration of the association (usually measured as a correlation) between a score on a given measure and performance on another assessment of the same or similar construct obtained at approximately the same time
*Construct validity**	Estimate of the degree to which an assessment measures the theoretical construct it claims to measure and to which inferences based on the assessment are relevant to the construct
Ecological validity	The degree to which the assessment uses a representative sample of tasks or behaviors in a context that is familiar enough for the child to recognize what is required
*Predictive validity**	The extent to which the measure's results are related to later functioning

* Malone et al. 2010

B. The Importance of Developing a Measure of Early Childhood Teachers' Use of Ongoing Assessment to Individualize Classroom Instruction

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, 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. But recent years have seen a rising interest in how teachers use ongoing assessment to adjust their teaching in order to best meet each child's needs. In fact, individualized teaching has been deemed a "best practice" in early education programs and is a requirement in the Head Start Performance Standards (*Federal Register* 2011).

Teachers have been using ongoing assessment in K–12 special education for more than two decades (Reschly and Ysseldyke 2002).[5] In K-12, ongoing assessments typically are not used to individualize instruction for all students; rather, teachers use ongoing assessment to identify students at academic risk and to pinpoint their weaknesses (Safer and Fleischman 2005; Shapiro 2008). Teachers use assessment data to make decisions about instructional grouping (Fuchs et al. 1992), to identify skill strengths and deficits (Foegen 2008; Foegen et al. 2007; Fuchs et al. 1991a; Whinnery and Stecker 1992), to screen children for potential school failure (Speece and Case 2001), and to determine children's eligibility for services (Buysse and Peisner-Feinberg 2013).

One common form of ongoing assessment in K–12 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). One of the most recent wide-scale applications of progress monitoring is response to intervention (RTI)—an approach to early intervention that includes the regular screening of all children throughout the year (Hamilton et al. 2009; National Association for the Education of Young Children et al. 2012; Buysse and Peisner-Feinberg 2013). In RTI systems, children who are falling behind receive intensive interventions as well as frequent monitoring to gauge the success of the interventions.

Several experimental studies have shown progress monitoring to be an important part of successful teacher support or professional development programs. In one experimental study, infants and toddlers whose home visitors used progress monitoring and received web-based guidance in making data-based intervention decisions, saw more growth in their communication skills than did those whose home visitors did not use progress monitoring (Buzhardt et al. 2011, 2010). In another study, researchers found that a professional development program with progress monitoring helped teachers enhance their instructional practices (for example, teachers altered their lesson plans to include language and literacy activities that showed their understanding of the scope and sequence of the curriculum). These changes led to improvements in children's language and literacy skills (Landry et al. 2011).

According to studies on elementary school children, teachers who use 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 (Fuchs and Fuchs

[5] For further discussion of the history of progress monitoring, see the literature review in Volume II (Akers et al. 2013).

2006). For example, studies show that ongoing assessment in reading (sometimes combined with guidance for individualized instruction) raises teachers' awareness of students' current levels of reading proficiency and improves the instructional decisions they make (Connor et al. 2009; Fuchs et al. 1984). The use of ongoing assessment data—often merged with other professional development supports, such as mentoring—is also linked to growth in literacy outcomes in preschool through first grade (Ball and Gettinger 2009; Landry et al. 2009; Wasik et al. 2009).

Head Start recognizes the importance of ongoing assessment to individualize instruction for young children. Over the past five years, the Office of Head Start (OHS) has elaborated on its vision for preschool child and family outcomes, strengthened its focus on program and classroom quality in its monitoring system, and developed tools to support ongoing assessment in daily practice. For example, in 2010 OHS released the Head Start Child Development and Early Learning Framework, a blueprint for achieving the program's child-specific goals through the alignment of curricular approaches, assessments, and professional development activities. Individual centers sponsored by OHS are also contributing to these efforts. For example, the National Center on Quality Teaching and Learning has published a Framework for Effective Everyday Practice. This framework likens Head Start's approach to school-readiness to a house—with a foundation of engaging interactions and environments, a first pillar of research-based curricula and teaching practices, a second pillar of ongoing assessment of child progress, and a roof representing highly individualized teaching and learning. Complementing these efforts is the Advisory Committee on Head Start Research and Evaluation, convened by the secretary of the U.S. Department of Health and Human Services, which recently advocated investing in evidence-based and data-informed practices across all domains of quality teaching and learning. In its final report, the committee shared its priorities, including "a need for additional guidance to programs on how to define and assess the progress of children toward school-readiness goals" and "how to achieve those goals through quality teaching and learning practices" (Advisory Committee on Head Start Research and Evaluation 2012).

But despite the importance of using assessment data to inform instruction in Head Start, there is little evidence on how early education teachers actually collect and use such data to tailor their teaching. Policymakers, practitioners, and researchers continue to see an urgent need for such evidence in the quest for better educational outcomes (Bambrick-Santoyo 2010; Black et al. 2003, 2004; Fuchs and Fuchs 2006; Hamilton et al. 2009; Marsh et al. 2006). The findings from this study could help fill this knowledge gap, particularly in Head Start, by defining the key aspects of the how teachers use ongoing assessment to individualize instruction and by providing guidance on measuring this process.

C. Literature Review

We began this project by conducting a literature review with two goals in mind. The first was to identify the critical areas we would need to address in our measure of early childhood teachers' use of ongoing assessment for individualization. The second was to find examples of how others have measured this process (Volume II of this report presents the literature review[6]). We conducted a library search targeting research related to early childhood education (which we defined as including children from birth through 3rd grade) and early childhood special education. Our search was limited to references from the past 10 years (2002–2012). In addition, members of the expert consultant group recommended studies for the literature review, including research that was not yet

[6] Akers et al. 2014

published. Together, the library search and the expert recommendations yielded 1,325 unduplicated references. A team of trained reviewers screened all references for relevance, identifying 173 references for review. The review team then extracted information from the references along key topics identified in consultation with ACF. Exhibit I.2 shows our main findings from the review.

Exhibit I.2. Key Findings from the Literature Review

- Ongoing assessment of child progress is increasingly a priority in early childhood classrooms, yet teachers' use of ongoing assessment has not been extensively researched.
- There is little rigorous evidence about the activities and supports that are critical for successfully using ongoing assessment to individualize instruction. The existing literature:
  - Describes the perceived best practices in using ongoing assessment for individualization as well as the range of ongoing assessment activities we are likely to see in early childhood settings
  - Does not provide guidance on how to determine whether these activities are well-implemented, nor does it describe the factors that affect teachers' abilities to implement the activities well
- We lack a solid research base on how to measure teachers' use of ongoing assessment for individualization.
  - Few studies measured teachers' implementation of ongoing assessments for individualization; of those who did, only some provided detailed information about the measure(s) used.
  - Across all components of implementation, we lack evidence linking measures of teachers' use of ongoing assessments for individualization to child outcomes.
  - A new measure would need to go beyond what is already in the literature by capturing an array of activities involved in the ongoing assessment process, considering implementation across a range of ongoing assessment tools, and capturing domains other than language and literacy. The measure would also need to apply to early childhood settings.

D. Roadmap to the Report

Based on findings from the literature review and input from experts and ACF, we developed conceptual frameworks to help us (1) define the key aspects of using ongoing child assessment to individualize instruction and (2) develop a measurement plan for examining teachers' implementation of this process. In Chapter II, we describe the conceptual frameworks for the two most common approaches to ongoing assessment in early childhood settings. In Chapter III, we describe the multi-method measure that we propose developing to examine how well teachers are using ongoing assessment to individualize instruction. Chapter IV concludes with the next steps for developing and pre-testing the measure.

This page is left intentionally blank

CHAPTER II. CONCEPTUAL FRAMEWORK

In this chapter, we present the key assumptions we made in developing two conceptual frameworks for this project. We then situate the frameworks within a theory of change—that is, a model of how ongoing assessment to individualize instruction leads to better outcomes for children—that depicts the inputs, activities, outputs, and outcomes. Next, we present the two primary approaches to ongoing assessment used in early childhood settings (general outcomes measures, or GOMs, and curriculum-embedded approaches) and their accompanying conceptual frameworks. We also describe several indicators of quality to be evaluated in each stage of the conceptual frameworks; these quality considerations serve as the foundation for our proposed measure of teachers' use of ongoing assessments to individualize instruction.

A. Assumptions Underlying the Conceptual Frameworks

A conceptual framework is a tentative theory that may explain a phenomenon being investigated (Maxwell 2011). The framework includes all the concepts, assumptions, expectations, and beliefs that support and inform the research being conducted on the phenomenon (Miles and Huberman 1994; Robson 2002). For this project, the frameworks are based on our review of the literature, discussed in Volume II of the report, and feedback from an expert consultant group.

We made the following assumptions when creating our frameworks:

- **Distal factors.** The frameworks include several distal factors, or factors that may be out of the immediate control of the teachers. These factors include the type of ongoing assessment tool selected by the early childhood program staff or the supports available to the teacher (for example, whether the teacher has enough planning time to review the assessment data or sufficient training on how to use the assessment tool).
- **Dynamic process.** The implementation of an ongoing assessment system is a dynamic process. As teachers conduct assessments, they become better at using the tools and interpreting the results. Similarly, as they modify their instruction based on the assessment data, they are better able to differentiate strategies that work from those that do not. In the frameworks, we assumed there would be a "spiraling" of teacher learning—each time a process was repeated, the teacher would acquire more expertise.
- **Teachers' knowledge and beliefs.** Teachers' pedagogical knowledge (including their knowledge of assessment) and beliefs are the lenses through which they view the assessment and individualization process. Because this is a dynamic process, teachers' knowledge and beliefs will likely change as they gain experience conducting assessments and making decisions based on the data.
- **Child outcomes.** Children can benefit from the ongoing processes of assessing, interpreting the data, making instructional decisions, and applying those decisions to improve classroom practices.

In the next section, we present a theory of change that illustrates how effective use of ongoing assessment to tailor instruction can promote student learning. This theory provides a wider context for the conceptual frameworks described in the remainder of the chapter.

B. A Theory of Change

Our measure would examine how well teachers use ongoing assessment data to (1) tailor their instruction for the entire group as well as for each child; (2) determine whether current instructional approaches are helping children become ready for school; and (3) establish the level of additional support or type of instructional modifications needed, if any. We can get a sense of the activities and supports that are important for successful use of ongoing assessment for individualization from the research literature, along with current practices in teacher education, professional development, and supervision. Figure II.1 shows some of the critical activities associated with ongoing assessment, the factors that affect the implementation of the activities (inputs), and how the activities lead to more refined instructional strategies and greater child progress (outcomes). The figure also shows several contextual factors likely to help or hinder the long-term outcomes.

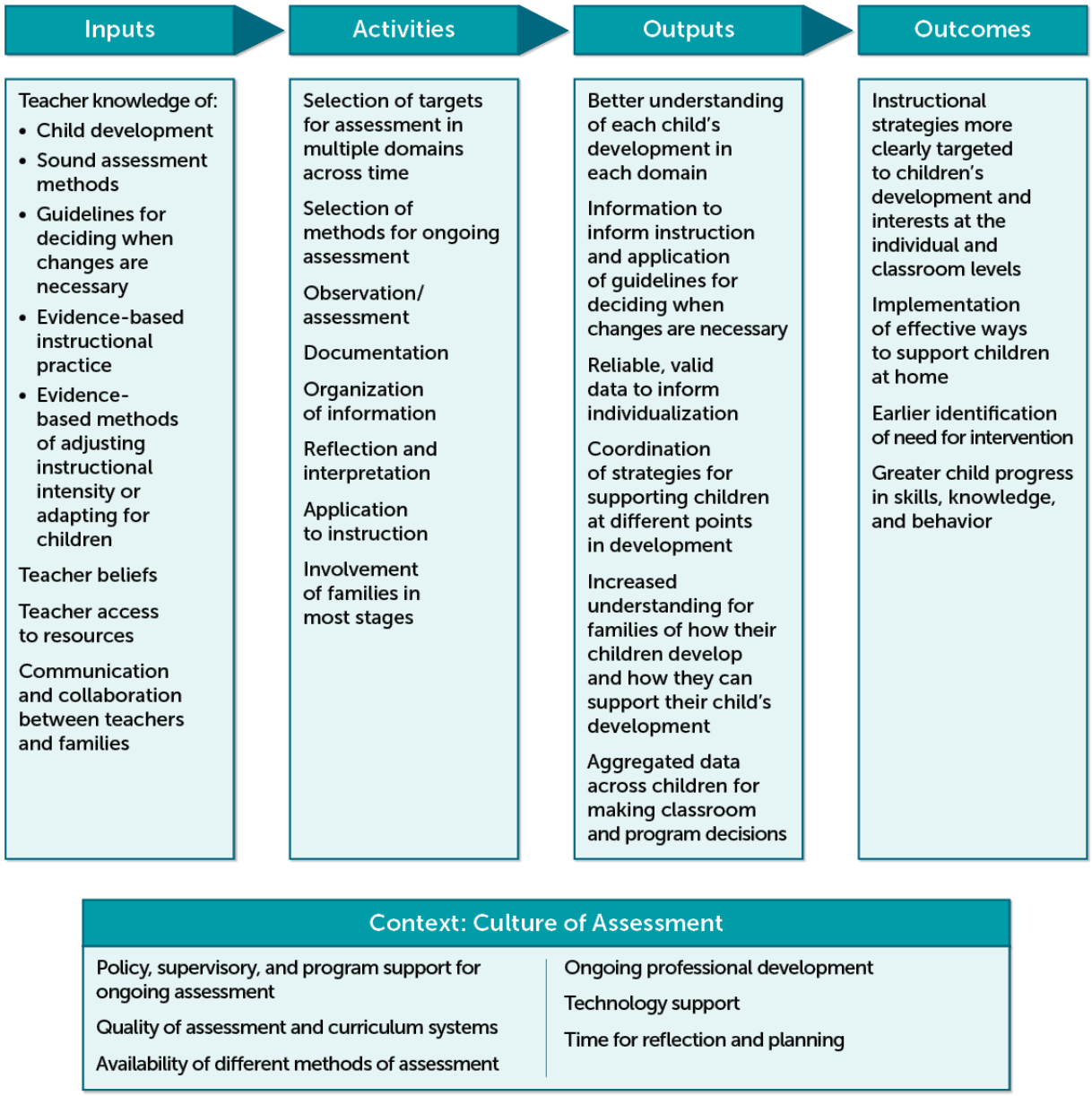
Inputs. Inputs are the resources that help make ongoing assessment possible. They include teachers' knowledge of and beliefs about child development, assessment, and evidence-based instruction. They also include teachers' access to various resources, such as developmentally appropriate benchmarks, reliable and valid measures, a curriculum with clear goals and sequenced objectives, and program support for the use of assessments. Communication with families is also a vital input, as it is important to incorporate their perspectives into the assessment process and encourage them to support children's development at home.

Activities. Activities are the iterative steps that make up the assessment and individualization process. The activities listed in Figure II.1 are explored in greater depth in the conceptual frameworks, and they constitute the constructs to be evaluated in a measure of teachers' use of ongoing assessment to individualize instruction.

Outputs. Ongoing assessments can lead to a number of outputs, including a better understanding among teachers of the children's interests, strengths, and challenges, both across domains and over time. This in turn enables teachers to better match their instruction to each child's strengths and needs, to keep parents updated on progress, and to communicate with parents about supporting children's development. Across children, this information can also be used to guide classroom instruction, and across classrooms, this information can be used to guide professional development and to understand program effectiveness.

Outcomes. The primary outcomes of the assessment process are instructional practices that promote learning and are well-suited to children's needs. But other outcomes are also possible, such as the earlier identification of children who are struggling and the implementation of better ways to support children's development both at home and in school.

Figure II.1. Assessing Teachers' Use of Ongoing Child Assessment to Individualize Instruction: Theory of Change



Context. It is important to note several contextual factors that may help or hinder teachers' use of ongoing assessment for individualization and affect the outputs and outcomes. These factors are part of the "culture of assessment" in an early childhood program setting and include policy and supervisory support for assessment, other program support (such as the availability of a teaching assistant), and the quality of the assessment system and curriculum used. Other contextual factors include the availability of different methods of assessment, ongoing professional development, technology support, and time for reflection and planning.

Logic model activities as basis for conceptual frameworks. The conceptual frameworks for this project expand the considerations within the "activities" column of the theory of change. In the remainder of this chapter, we present two conceptual frameworks, one for each of the most

commonly used approaches to assessment in early childhood programs: GOMs and curriculum-embedded approaches. These approaches offer different types and levels of supports for collecting, organizing, interpreting, and using data as well as for selecting what to measure. GOMs provide more support for teachers than curriculum-embedded approaches, which require more decision-making on the part of the teacher. Thus, additional indicators of quality need to be examined for the curriculum-embedded approaches to capture the quality of teachers' decision-making.

Next, we describe each approach and its conceptual framework. We begin with the GOM approach and then present the curriculum-embedded approach, elaborating on the indicators of quality needed for successful use. These indicators of quality can inform the development of a measure. We conclude with a justification for developing a measure that examines quality in the more frequently used curriculum-embedded approaches.

C. The GOM Approach

In the GOM approach, teachers use a brief measure with strong evidence of reliability and validity to conduct frequent, standard assessments of children's progress toward a long-term goal. Central to this approach is the repeated measurement of a few key skills—which represent the entire skill set required to achieve a given goal—rather than measuring the full skill set. Proficiency in these few skills predict the child's later success with the entire skill set; for example, a GOM might require a child to name pictures fluently to predict the breadth of the child's vocabulary and background knowledge. A child's increasing proficiency on a GOM is indicated by improved performance on those key tasks repeatedly as measured over time (for example, the ability to name more pictures in a set). Children's performance using GOMs can be measured as little as three times per year or as often as once per week (Jenkins et al. 2009), and the pictures, questions, or activities used to obtain these performance samples typically range from one to five minutes, depending on the outcome being measured.

GOMs have many applications. Because they are brief, they can be used to screen all children to identify those at risk for delays in certain areas such as early literacy, vocabulary, and social and emotional development. The criteria for identifying children with a given delay are often based on national norms (such as average performance for a national sample of children) but can also be based on local norms (such as average performance in a given classroom). GOMs can also be used to determine whether instructional changes led to changes in a child's growth rates. Thus, GOMs have become a key component in many RTI models. GOMs have been used most extensively to measure growth in reading, writing, and math in the elementary grades, but researchers have developed GOMs for younger children as well. Specifically, GOMs have also been developed for infants and toddlers in the areas of early communication, social development, movement, and problem-solving (Carta et al. 2010). For preschoolers, GOMs have been developed in the areas of early literacy, language, and math (Missall et al. 2008; Hojnoski et al. 2009).

GOMs have been in use for more than 20 years. One of the most widely used is the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) for children in kindergarten and first grade (Good et al. 2001; Kaminski and Good 1996). DIBELS is a set of measures of early literacy that are sequenced and aligned with instruction in early elementary school; they are currently administered yearly to more than six million students across the nation (University of Oregon Center on Teaching and Learning 2012). More recently developed GOMs include the Individual Growth and Development Indicators (IGDIs), targeted to younger children. To date, IGDIs have been developed to assess the progress of infants and toddlers in language, social development, movement, and cognitive outcomes (Carta et al. 2010, 2004; Greenwood et al. 2002). IGDIs developed for

children ages 3 to 5 include Picture Naming (a measure of expressive language) and a suite of phonological awareness measures such as Rhyming, Alliteration, and Segment Blending (McConnell et al. 2002). Research indicates that the IGDIs are reliable and valid as well as able to detect growth by age, disability status, and level of risk (Carta et al. 2010; Carta et al. 2004; Greenwood et al. 2002).

Research on the use of GOMs in elementary grades suggests the assessments are beneficial in several ways. For example, teachers using GOMs are better able to identify students who need additional or different forms of support.[7] These teachers also tend to design stronger, more effective instructional programs than teachers who do not use GOMs. In addition, students in classrooms where GOMs are used achieve better outcomes (Fuchs and Fuchs 2006; Good et al. 2001). But GOMs have also been shown to be viable in grades earlier than elementary. For example, studies have shown the reliability, validity, and feasibility of the GOMS for infants and toddlers, including both longitudinal and cross-sectional studies illustrating the GOM's ability to detect growth over time (Carta et al. 2004; Greenwood et al. 2002; Greenwood et al. 2006; Luze et al. 2001). Single case studies have also shown GOMs' ability to detect growth over time in response to short-term early interventions with infants and toddlers in Early Head Start, child care, and early intervention programs (Greenwood et al. 2003; Harjusola-Webb 2006).

Strengths

Every approach to gathering information about child progress has strengths and weaknesses. GOMs are distinguished by five specific strengths. First, they are brief and easy to complete, requiring little staff expertise or training, and are therefore suitable for wide-scale and frequent use. Second, they allow users to directly assess child performance on common tasks using a standard metric, allowing the assessment of growth. Third, many GOMs appear to be effective with diverse groups and can apply to children with a range of abilities. Fourth, GOMs can detect changes over short periods, allowing for fine-grained assessment of child progress. Fifth, the GOMs used in RTI adhere to high standards of reliability and validity, including concurrent and construct validity as well as predictive validity (Deno 1997; McConnell et al. 2002).

Weaknesses

GOMs also have several weaknesses. For example, they are traditionally used to help teachers identify and individualize lessons for children who are struggling; however, Head Start requires that teachers individualize for *all* children, including those who are already getting appropriate support based on their universal screening results (for example, children who are doing exceptionally well should also be receiving challenging individualized instruction). And although GOMs help pinpoint children who may be falling behind, these measures do not necessarily help a teacher determine *what* to teach them. Low performance on a GOM indicates that the teacher should change his or her instruction in a particular area, but there is no guidance on specific changes. Instead, the teacher must make decisions about how to change instruction based on her own understanding of the child's performance on skills outlined in the curriculum.

[7] When interpreting these findings for the purpose of this study, it is important to note that elementary school teachers typically hold advanced degrees and have access to more curricular support than preschool teachers.

These measures have an additional disadvantage when used in early childhood settings: GOMs for young children are only available for a limited number of skills and do not address all the domains identified for assessment in the Head Start Child Outcomes Framework. As a result, teachers with limited knowledge about development and assessment may mistakenly view these few skills as the entirety of what children need to know. Careful attention is therefore needed to understand whether and how teachers use GOMs to guide instruction and individualize for a wider skill set.

D. Conceptual Framework for the GOM Approach

When GOMs are used to individualize instruction for young children, the assessments must be quick and easy to use, suitable for frequent use, and able to detect clear growth toward a meaningful outcome. The assessments should also be able to show the child's skill level and rate of progress compared with national or local norms.

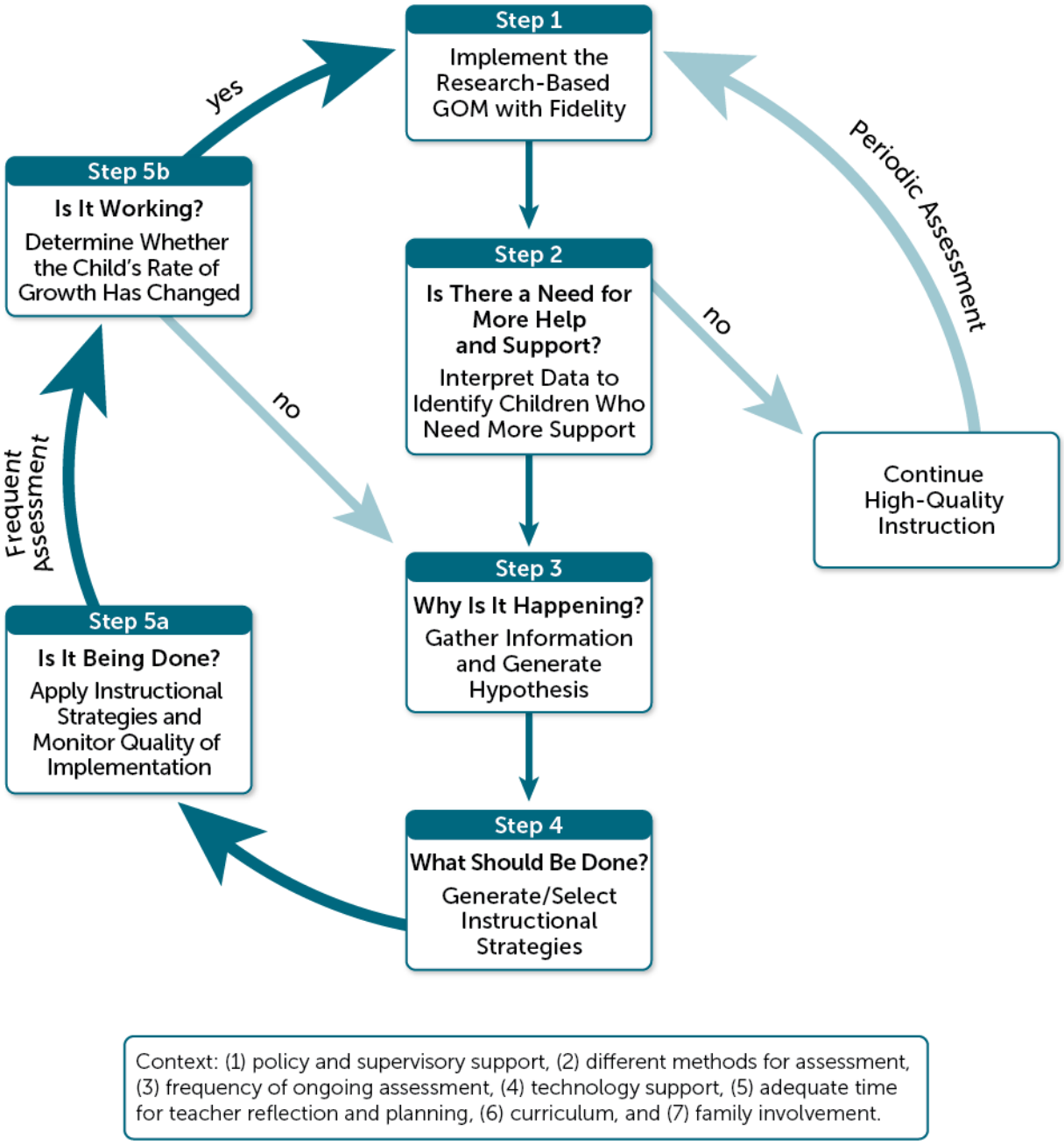
Programs will generally select the assessments and provide them to teachers (to assess all children in a class) or to home visitors (to assess all children in a caseload). In either case, the programs should make sure the selected GOMs meet all traditional standards of measurement quality (reliability and validity) as well as the specific quality standards for ongoing assessment (Fuchs and Fuchs 2006). The reliability standard requires that the data generated by the assessment be accurate and comparable across a given child's data and across data for all children, whereas the validity standard requires that the assessment actually examines the skill or domain of interest (American Educational Research Association 1999). GOMs also involve the repeated measurement of the same skill; for this reason, they must include alternate forms of each test, and all the tests must produce equivalent score estimates for individual children. For example, suppose an assessment includes an expressive vocabulary test that requires naming a set of pictures. The measure would need to have multiple versions of this test that are equal in difficulty in order to be considered a reliable GOM.

The GOM approach to individualization often used in preschool (sometimes called the problem-solving approach) follows a decision-making model based on the work of Tilly (2002) and Deno (1997). In this model, GOMs are administered frequently (at least quarterly) to all children. After implementing the GOM with fidelity, teachers follow a decision-making process including questions at each step to determine how to support an individual child's learning.

Figure II.2 shows the steps involved in using GOMs for ongoing assessment. As shown in the figure, GOMs are repeated frequently throughout the year, with some steps occurring more frequently if the instructional changes do not appear to be working for a particular child. It is important to note that a number of factors can affect how the GOMs are used and whether they work, including the program context (such as supervisory support for assessment and the quality of the assessment systems used), teacher beliefs about the importance of assessments, communication with families, and teacher knowledge of evidence-based or professionally recommended practices.

In the next section, we discuss the steps followed in the GOM approach and the indicators of quality for each step.

Figure II.2. Conceptual Framework for the GOM Approach



Step 1: Implementing the GOM Assessment with Fidelity

Using GOMs with fidelity means that the teacher is faithful to the GOM protocol for what, when, how, and whom to assess. For example, the centerpiece of the GOM approach involves screening all children on each of the GOMs selected by the program (for example, communication, motor skills, and social and emotional development). To be implemented with fidelity, this screening should be ongoing for all children, at least on a quarterly basis, and more frequently for at-risk children or children with disabilities. GOMs also generally have a precise administrative protocol that should be followed to ensure that data accurately reflect each child's performance. In addition, teachers who use the GOMs should be certified to do so. They must follow the administrative guidelines, such as using the standard set of materials; ensuring that the child is comfortable in the assessment situation; and adhere to any time limits for the assessment's administration.

When GOMs are used across an entire program, indicators of quality implementation might include (1) evidence that assessments are administered by certified staff, (2) frequent reliability checks between teachers, (3) adherence to established timing guidelines, and (4) consistent use of an organized system for recording data.

Step 2: Interpreting the Data (Is There a Need for More Help and Support?)

When a child needs additional support, his or her score on a GOM will be significantly below the norm for the child's age. If the score is well below the norm (more than one standard deviation below the normative mean), the researcher will often screen the child again relatively soon to confirm that the original score is valid. Indicators that this step has been implemented correctly include accuracy and fidelity in entering data, reading graphs, and following the recommended guidelines for deciding when children need more help and support.

Step 3: Generating Hypotheses (Why Is It Happening?)

After confirming that the child needs additional support, teachers should find out why. They may need to collaborate with a home visitor or other service provider who works with the child's caregivers to determine why a child may be experiencing a delay. For example, a child's recent transition to a new Head Start classroom may have lowered his or her score on a GOM measuring social or communication outcomes. This process helps to identify evidence-based instructional strategies appropriate to meet the child's needs.[8]

This step includes four quality indicators, the first of which is the extent to which the teacher gathered data from a variety of sources (including comparison to benchmarks or guidelines indicated in interpreting the assessment data). Other data sources may include other assessments, classroom performance records, and health records. The second indicator is the extent to which the teacher considered how the child is responding to the core program used with all children in the class. The third is how well the teacher accounted for the child's home, cultural, and linguistic diversity. The fourth is whether the teacher considered any learning difficulties noted in the past. Working alone or with a team (for example, with a home visitor, other service provider, and parent), teachers use this information to deduce why a child might be struggling or what type of support is needed.

[8] If evidence-based strategies are not available, professionally recommended strategies should be identified.

Step 4: Generating/Selecting Instructional Strategies (What Should Be Done?)

Once it is known that a child is struggling, the teacher can use one of three approaches to develop an evidence-based intervention strategy. In the first option, the problem-solving approach, the teacher uses data to determine the type and degree of assistance the child needs. In the second option, the standard protocol approach, the teacher consults the materials that accompany the GOM; these materials provide recommendations for specific standardized, research-based strategies to help children who have certain types of difficulties. For example, for a child who has a limited vocabulary, the materials might recommend an early literacy/language intervention that includes small-group intensive reading instruction three times a week. This approach helps to ensure fidelity of strategy implementation. A third approach, called Making Online Decisions (MOD), combines elements of the first two approaches and was developed for use with infants and toddlers. Via a web-based system, MOD guides practitioners through multiple strategies that follow a standard protocol based on the child's age and performance in a given area (Buzhardt et al. 2010, 2011). With this approach, the practitioner (usually a home visitor) collaborates with family members to select strategies that parents and caregivers can use with the child at home. The practitioner continues to evaluate the child on a regular basis and uses the same web-based system to enter data and generate data displays (typically charts or graphs), which illustrate whether the child is benefiting from the intervention.

Regardless of the approach used, the quality indicators for this step are the same. The quality of implementation is based on the match between the child's characteristics (particularly data about the child's functioning and progress) and the selected evidence-based strategies.

Step 5: Applying Instructional Strategies, Assessing Implementation, and Determining Whether the Child's Rate of Growth Has Changed (Is It Being Done? Is It Working?)

Whether the intervention is customized for a child or is a standard strategy, some planning is needed to implement it. The team responsible for carrying it out (such as parents, teachers, and home visitors) must plan when, where, and for how long the strategy will be used and who will implement each part of it. They must also decide how the strategy will be implemented and what the goals are (for example, an increased rate of growth on a specific GOM, such as increased rates of gesturing to communicate wants and needs).

During implementation, fidelity is monitored (*Step 5A: Is It Being Done?*), and if the child is making less-than-expected progress, he or she is assessed more frequently to track progress towards the targeted outcome (*Step 5B: Is It Working?*). At this point, frequent assessments help the team make decisions about whether to change strategies. Based on the child's performance and growth trajectories, the strategy might be kept in place or may be modified (for example, increasing the frequency, duration, or intensity). Alternatively, a new strategy may be used, or in a best-case scenario, the strategy might be phased out if the child is showing progress (that is, an increased growth rate).

To gauge the success of the strategy, we would look at the fidelity of the implementation, including whether the child received the intervention with an adequate frequency and duration and how engaged the child was in the intervention strategy. The instructional team would keep assessing the child using one or more GOMs to see whether the child is improving (that is, whether the rate of growth changes in response to the intervention). If so, the team would assume that the intervention fit the child's needs. If not, the team would use the procedures described above to make changes.

E. The Curriculum-Embedded Approach

The most commonly used approaches for assessing the progress of children in early care and education are curriculum-embedded approaches. Curriculum-embedded approaches are commonly used in preschool classroom assessment to inform daily instruction and interventions, identify children's strengths and difficulties, and assess progress across time (Early Learning Standards Task Force and Kindergarten Assessment Work Group Pennsylvania BUILD Initiative 2005). These approaches help teachers assess children's progress relative to early learning standards and the skills and knowledge taught in a specific curriculum. Teachers often conduct these assessments as they are teaching their classroom's curriculum. The assessment tasks are intended to be authentic in context; that is, they are "opportunities created for children that reflect typical experiences, rather than discrete, isolated tasks that are irrelevant to the child's daily life (Pretti-Frontczack et al. forthcoming). Some of these assessments (known as "curriculum-based" assessments) are created by the curriculum developers to align very closely with the material being taught, whereas others (known as "curriculum-embedded" assessments) are derived from national standards and developmental expectations.

Teachers using a curriculum-embedded assessment system will typically assess their children's performance in relation to indicators on rubrics provided by the system. The rubrics specify different levels of performance building up toward end-of-year goals but often provide no guidance regarding children's expected progress throughout the year. In addition, although the tasks being assessed are embedded within daily activities and aligned with curriculum goals, they are not standardized; therefore, teachers collect assessment data from many different sources. Sources may include observation recording forms, worksheets, portfolios, and standardized assessments that are aligned with the curriculum.

Teachers assess each child based on his or her performance on a sequence of increasingly difficult skills that may be tracked with separate, though related, tasks or assessments over the course of the year. Evidence of the child's performance and progress is often obtained from several sources, usually within authentic contexts (Bagnato et al. 2009; Bennett 2011; Meisels and Atkins-Burnett 1999). This sets curriculum-embedded approaches apart from GOMs, which gauge a child's proficiency in only a few critical skills and are not typically delivered in authentic contexts.

A curriculum-embedded approach can be used formatively to assess a child's progress, rather than just a child's current mastery of the knowledge, skill, or behavior in question.[9] When used formatively, teachers ask a child who has already demonstrated a skill in one context to demonstrate the same skill in a new context. Teachers also assess the child's current learning, taking into account the contexts and supports for learning.

Evidence of the reliability of curriculum-embedded assessments includes internal consistency, rater agreement (Lambert et al. 2010; Meisels et al. 2001), and test-retest reliability (Landry et al. 2009), although not all curriculum-embedded assessments evaluate the psychometric integrity of the assessments. In addition, rater agreement has been examined only under somewhat atypical conditions—that is, studies in which the volunteer teachers received a great deal of professional

[9] In special education, formative assessments are used to provide information about how students are responding to instruction and whether they are making progress toward goals and objectives (Hamilton et al. 2009).

development (Meisels et al. 2001) and studies in which all teachers rated the same video recording of a child who was unfamiliar to them (Lambert et al. 2010).

Only a few curriculum-embedded assessments in early childhood settings have evidence for concurrent and predictive validity evidence (Meisels et al. 2001, 2003). The few studies addressing validity are plagued by problems with the study design, such as possible bias caused by teachers volunteering for the studies and the presence of other innovations in the schools that might skew the results. We see more positive child outcomes in certain kinds of studies using curriculum-embedded approaches in early childhood, including case studies (Bambrick-Santoyo 2010), observational studies (Bodrova and Leong 2001; Meisels et al. 2003), and mixed-method studies (Bambrick-Santoyo 2010; Goertz et al. 2009), as well as more rigorous randomized-controlled studies (Al Otaiba et al. 2011; Landry et al. 2009). Most of these studies focused on outcomes in the literacy domain, but some limited evidence of positive child outcomes was also found for several other learning domains in early childhood, elementary, and secondary schools (Bambrick-Santoyo 2010; Goertz et al. 2009; Meisels et al. 2003).

Reviews of research show that one form of curriculum-embedded assessment—formative assessment—has positive effects in elementary and secondary schools (Black and Wiliam 1998a, 1998b). However, these studies are disparate, measuring different aspects of the assessment and individualization process. Individual studies have equivocal findings, with some negative outcomes attributed to the differences in student characteristics, the type of tasks, the domains, the part of the process examined, and the high need for teacher knowledge in implementation (Bennett 2011). One conclusion we can draw is that effective use of formative assessments requires valid, meaningful assessment; evidence-based inferences; and high-quality, evidence-based adaptation of the curriculum (Bennett 2011).

Strengths

The primary strengths of curriculum-embedded approaches are strong alignment with the curriculum, the integration of the assessment with classroom activities, and the focus on careful observation of children and reflection on their performance. By aligning the assessment system with the curriculum, teachers can use the curriculum to identify what the next step for the child should be and can draw on the curriculum for instructional activities and teaching strategies. In fact, some curriculum-based assessments use hand-held devices or other technology to continually provide teachers with such strategies (Al Otaiba et al. 2011; Bodrova and Leong 2001; Landry et al. 2009). In addition, many rubrics include behavioral descriptions for children at each skill level, which can help teachers identify the next instructional step for a given child.

Weaknesses

Despite these strengths, curriculum-embedded approaches have several weaknesses, including limited evidence of reliability and validity, a high level of subjectivity in the teacher ratings, and a strong reliance on teacher knowledge. Though easy to link to classroom activities and instruction, curriculum-embedded assessments require more teacher knowledge of assessments and skill in conducting them. GOMs, on the other hand, provide specific assessment targets and explicit guidelines for interpreting the results.

With the autonomy of the curriculum-embedded approaches comes the risk of poor implementation: teachers may not have the knowledge needed to select good assessment targets or correctly interpret child behavior. For example, it is important to continue to assess mastery of prior

as well as current objectives (skills, knowledge, behavior) to ensure that children retain what they have learned (Bennett 2011). However, some teachers may focus too much on current learning, or they may not recognize when a child is falling behind or know how to help the child. To successfully use a curriculum-embedded approach, teachers need more professional training (Heidemann et al. 2005; Landry et al. 2009, 2011) and/or technology support than they would if they were using GOMs.

F. Conceptual Framework for the Curriculum-Embedded Approach

In this section, we describe how curriculum-embedded approaches are used to inform and tailor instruction. Figure II.3 shows the stages of this process and some relevant contextual factors. For each stage, we list indicators of quality that should be considered when developing a measure of how well teachers use a curriculum-embedded approach. We selected these indicators because they are critical for a high-quality implementation (in other words, they are meaningful), they can be changed (for example, via professional development), and they generalize across early childhood settings and domains of child development.

The Importance of Context

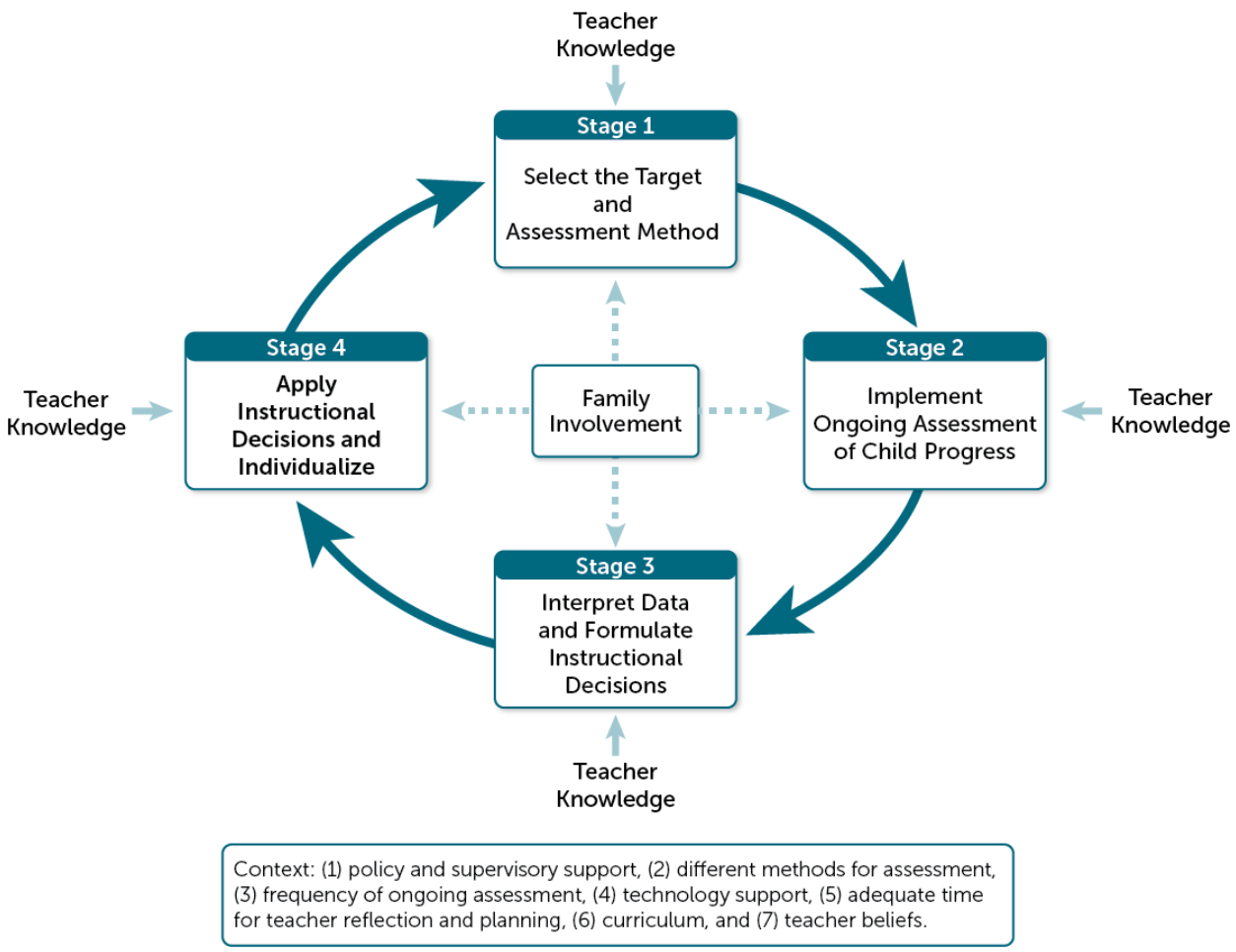
Many curriculum-embedded systems offer support to teachers to conduct ongoing assessments and individualize. Recent innovations include systematic approaches to individualization, with standard elements that help teachers gather information on a child, organize and interpret it, compare it to national or local norms, and use it to modify their instruction. (For example, based on a child's results, a system may recommend specific lessons to teach using specific strategies.) However, early childhood programs are diverse, and thus the assessment systems they use may not offer all teachers support in all aspects of the process. Even when the support is there, teachers may not be able to use the assessment system with fidelity. Therefore, in our conceptual framework for curriculum-embedded approaches, we did not assume that technological supports and coaching are available for teachers. But the context in which assessment occurs—including the available supports—does affect the quality of implementation, so it is important to consider key contextual factors that will help or hinder ongoing assessment. These factors include the following:

- Policy and supervisory support for conducting ongoing assessments. This includes the availability of training, coaching, technology (such as access to personal digital assistants or computers), or even supports as minimal as clipboards and materials for organizing children's work. Policy sets the stage for the other aspects of context.
- Policy requirements and support for *frequent* assessments of child progress. This includes a schedule for assessing children as well as for reviewing and reflecting on progress. It is difficult to use information to tailor instruction if teachers reflect on it only two or three times a year.
- Availability of adequate time for reflection and planning
- The culture of data use and collaboration between teachers in assessing and interpreting the data
- Access to information about instructional strategies that are evidence-based or professionally recommended and aligned with the curriculum

- Availability of professional development opportunities, coaching, or consultation geared toward helping teachers deliver high-quality, evidence-based instruction that is tailored for each child

Figure II.3 shows the conceptual framework for a curriculum-embedded approach. The stages are iterative, beginning with (1) selecting the assessment target and method, followed by (2) conducting the assessment, (3) interpreting the data (including developing a hypothesis and making instructional decisions), and (4) applying the decisions. One clear difference in this model versus the GOM model is the frequent selection of the assessment target and method—that is, what skills or behavior to assess and how to assess them. Below, we discuss each stage and the indicators of quality implementation at each stage.

Figure II.3. Conceptual Framework for Curriculum-Embedded Approaches



Stage 1: Selecting the Target and Assessment Method

The assessment system is usually selected by program staff rather than by teachers. However, teachers have some autonomy in selecting the assessment target (the skill, knowledge, or behavior to be assessed) and the assessment method (how that skill, knowledge, or behavior will be assessed)—although both are also affected by the assessment system. The selected assessment targets should represent the skills or knowledge a child would need to meet the specified end-of-year goals. Teachers should collect, interpret, and reflect upon the data and make instructional adjustments

throughout the year, frequently enough to monitor the expected changes in performance and to track progress. Their decisions will play a larger role in this process than with the GOM method. Also, unlike GOMs—which define a small set of assessment objectives—curriculum-embedded approaches typically rely on rubrics that define a larger set of objectives by describing stages in development of an instructional goal and sometimes propose different kinds of evidence that can be collected to assess progress toward the goal.

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, and able to be generalized; whether the data collection method is valid; and whether data can be collected frequently. Below, we discuss these aspects in detail:

- The teacher selects assessment targets that are valid, generalizable, and able to be assessed among all children.
  - The teacher selects an assessment target that focuses on a key indicator of development and is linked to meaningful outcomes—that is, a skill, knowledge, or behavior that a child needs in order to be successful now or in the future.
  - The teacher selects an assessment target that is developmentally appropriate (for example, looking at different ways to make a set of five using concrete objects—two buttons and three buttons or four buttons and one button—rather than asking children to solve written equations).
  - To meet Head Start policy regulations, the teacher selects an assessment target that addresses key school-readiness domains (language and literacy development, cognition and general knowledge, approaches to learning, physical well-being and motor development, and social and emotional development).
  - The teacher selects an assessment target in an area where children of this age typically make progress within the program year (that is, change is expected in the current year). In other words, the evidence suggests that the targeted area can be changed with instruction or intervention, rather than by maturation only (for example, physical height is not amenable to instruction).
  - The teacher selects an assessment target that examines areas taught in the classroom curriculum (that is, the assessment is aligned with the curriculum).
  - The teacher selects an assessment target that is defined and measured by observable behaviors. The behavior may be observed through direct assessment (for example, asking the child how many items are in a set) or indirect evidence (for example, assessing a child's approach to learning by observing how long the child persists at a task).
  - The teacher selects an assessment target that can be assessed universally; in other words, all children, despite any disabilities, have a way to demonstrate progress on the assessment target.
  - The teacher selects an assessment targets that is generalizable; that is, the skill, knowledge, or behavior can be demonstrated in a variety of settings.

- The teacher selects an assessment method that is valid and feasible.
  - The teacher selects an assessment method that is aligned with the target. For example, if the teacher wants to assess the child's persistence, she should observe more than familiar tasks that the child finds easy. Observing the child in a challenging task would provide the teacher with a stronger measure of persistence.
  - The teacher selects an observation/assessment method that is valid for, or accessible to, the child. For example, if the teacher introduces the names of categories to a dual-language learner only in English, the child should be allowed to respond in English even if the question is asked in Spanish. When selecting the method or task, the teacher should keep children's differences in mind, including differences in culture (for example, whether the child eats with a fork or with chopsticks), temperament (for example, whether the child will demonstrate what she knows and can do only in small groups or individually), and ability (for example, whether the child needs modifications, such as special seating, to participate in an activity).
  - The teacher selects an observation/assessment method that is ecologically valid. That is, the tasks should be something that children would be expected to do, rather than something so unusual that children may not understand what to do. If children do not understand the task, they may not demonstrate knowledge or skills that they have. Assessing within a natural context can be more efficient for teachers—for example, it may be more efficient to observe whether a child can zip a coat when getting ready to go outside rather than during a small-group activity.
  - The teacher has an efficient means of collecting the data. If the data are too onerous to collect, teachers will be less likely to collect data often enough to assess progress.

- The teacher's schedule for collecting ongoing assessment data supports reliable and valid interpretation of child progress.
  - The teacher's schedule for collecting information about each domain allows him to review progress quarterly (with a minimum of three observations for a given assessment target). The teacher collects some data each week and collects data more frequently on the areas currently being taught.
  - The teacher's schedule for data collection matches the expectations for progress. In other words, the teacher collects data frequently enough to know when children need more or less support or challenge. (It is easier for teachers to collect and use ongoing assessment data if they have an assessment plan; however, given the nature of individualization, the plan may be dynamic.)
  - The teacher considers whether assessing some areas of learning at different points in the day (or week) may lead to different conclusions about child progress. For example, if the teacher assesses knowledge of letter names right after a group review of letters, he may find different results than if the children are assessed as they arrive on a Monday morning. Thus, assessments administered in various contexts could yield various patterns of change over time. Teachers may begin administering assessments within one consistent

context and then, when children are successful in that context, branch out to other contexts.

Stage 2: Implementing Ongoing Assessment

Given that teachers typically deliver these assessments during instructional activities, they should prioritize efficient assessment to maximize instructional time. The assessments should be ecologically valid. Teachers also need to document child progress 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. We discuss these criteria in detail below:

- The teacher's implementation of the assessment is valid.
  - The teacher implements the selected method in a way that is valid and can be replicated (see above).
  - The teacher ensures that the child understands the directions for the task (or assesses whether the child understands the directions).
  - The teacher can devote adequate attention to the child during the assessment—that is, the teacher is not distracted by other children or duties when observing a child's behavior.
- The teacher's documentation is reliable, valid, complete, and efficient.
  - The teacher is consistent in what and how the observations are documented, both within his or her own documentation and with other teachers' documentation.
  - The teacher's documentation is objective, stating what happens rather than making judgments.
  - The teacher notes important information about the child's behavior, the task, and the context of the assessment, including the date; time; group size; type of activity; presence of adult, peer, and/or environmental support; use of prompts; and any other adults taking part in the assessment.
  - The teacher uses an efficient documentation method, minimizing both the burden on himself and the time taken away from instruction. Teachers are likely to collect data more often if they have efficient means of doing so. 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 the responses.

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. Each child's data, combined with other available data (such as information about instructional activities, peers' performance, national benchmarks, and family input), would

help the teacher identify the child's strengths, weaknesses, interests, and learning differences and then select the best way to support the child's continued progress. The process of interpreting data and making instructional decisions may be conducted in teams with the support of other teachers, coaches, consultants, and family members.

Indicators of quality to be measured at this stage include how well the teacher has organized the assessment data and whether his interpretations are evidence-based and consider alternative hypotheses. These criteria are described in detail below:

- The teacher organizes information in a way that is efficient and supports valid interpretation.
  - The teacher's organization makes it easy to examine change over time in specific skills, knowledge, and behavior (facilitating examination of data collected from multiple sources using multiple methods across multiple time points).
  - The teacher's organization makes it easy to understand not only a child's performance compared with a benchmark, but also her strengths and interests.
  - If the assessment system organizes the data, the teacher uses the system with fidelity—for example, he enters data into the system correctly and promptly. The system may align with an integrated assessment plan that determines what data the teacher should gather, how, when, and for whom.
  - The teacher's organization makes the data easy to communicate to parents.
  - The teacher's method of organization imposes minimal additional burden. If the teacher is organizing the data, she does so efficiently.
- The teacher's interpretation of the data is valid and reliable.[10]
  - The teacher's interpretation is evidence-based:
    - The teacher draws on information from multiple time points, sources, and methods of assessment rather than from a single response.
    - The teacher recognizes that child performance is context-specific; in other words, it is related to the level of environmental or instructional support provided.
    - The teacher considers evidence-based (or professionally recommended) strategies to support child progress.
    - To evaluate progress, the teacher uses research-based or professionally recommended guidelines with an evidence base.
  - The teacher considers children's strengths, needs, and interests.
  - The teacher considers whether there is enough evidence to determine if the child is making expected progress over time, both in relation to how other children are

[10] This process is similar to steps 2 and 3 of the GOM approach.

doing (that is, in normative terms) as well as in relation to what the teacher would like the child to learn (that is, criterion-based terms), and identifies children who are not doing so.

- When appropriate, the teacher considers context and whether a child's behavior has been generalized.
- The teacher considers (and perhaps tests) alternative explanations for a child's behavior, regardless of whether the behavior was positive or problematic. For at least some assessment targets, the teacher develops hypotheses to be tested at the next instructional opportunity—for example, "Could she do the task if I provide more visual cues?" or "Could he do this in a different context?"
- The teacher involves other stakeholders, such as parents and 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.
- The teacher bases her interpretation on enough evidence that other early childhood professionals would interpret the data in a similar way.

Stage 4: Applying Instructional Decisions and Individualizing

Individualization is not only important for maximizing a child's progress but is a requirement of the Head Start Performance Standards ("Head Start Performance Standards" 2011). In terms of making instructional decisions, individualization involves selecting and using high-quality, research-based instructional strategies (or a professionally recommended instructional guideline with an evidence base) that are tailored for individual children and reflect what the teacher has observed about each child. To evaluate progress, the teacher uses research-based or professionally recommended guidelines with an evidence base.

The indicators of quality to consider at this stage include whether the teacher uses evidence-based strategies, differentiates instruction using a variety of approaches, and uses instructional strategies that build on children's strengths and interests.[11] We discuss these in detail below.

- The teacher implements evidence-based approaches that are responsive to the data with fidelity.
  - The teacher's approach is systematic and reflects the data (and any progress the children make).
  - The teacher's instructional approach is implemented with fidelity, the content is correct, and the level of rigor meets each child's needs.
  - 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.

[11] The processes here are similar to steps 4, 5a, and 5b of the GOM approach.

- The teacher differentiates instruction, where appropriate.
  - The teacher may use flexible, purposive small groups (grouping children in different ways to address various learning needs and strengths).
  - The teacher may individualize instruction within a group, offering different questions or different levels of prompts to each child.
  - The teacher may offer more opportunities for a child to practice a particular skill throughout the day.
- The teacher uses varied or individualized approaches to meet the needs of children.
  - The teacher's approaches may include using small groups or more intensive practice.
  - The teacher may adapt and modify activities for different children.
  - The teacher may use different prompting strategies.
  - The teacher may use 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.
- The classroom instructional team members share their knowledge about each child's goals and instructional strategies. Team members may include the lead teacher, assistant teachers, and any specialists involved in the classroom.

Teacher and Family Factors Affecting Most Stages

Unlike GOMs, which typically give teachers explicit guidance, curriculum-embedded approaches usually require teachers to make their own decisions about data collection, documentation, interpretation, and application. This means that the opinions and beliefs of teachers (and any other decision makers) have a greater effect on the overall process. Accordingly, the curriculum-embedded conceptual framework identifies two factors with implications for the entire process: (1) teachers' knowledge and beliefs about assessment, instruction, and children's development and (2) family involvement in the process of ongoing assessment.

Teachers' beliefs are a major factor in how they assess children and individualize instruction. For example, teachers who believe that using data leads to better outcomes for children are more likely to collect and use data. Knowledge of child development and pedagogical knowledge also play a key role. If an assessment system does not provide enough information on how to assess a child, for example, teachers with a greater understanding of child development will be better able than less-knowledgeable teachers to select appropriate instructional targets, determine valid methods for assessing and documenting, and interpret the data and changes over time. Teachers with a solid grasp of pedagogy will be 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. Such teachers will also have the knowledge needed to explain the process and results to families and engage families in the process.

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

G. A Focus on the Curriculum-Embedded Approach

The remainder of this report focuses on creating a measure of teachers' use of curriculum-embedded approaches. We concentrate on this type of approach because it is (1) more common in early childhood settings than GOMs; (2) more demanding for a teacher to implement (that is, it requires more teacher knowledge and assessment skills); and (3) traditionally more comprehensive, given that it focuses on many domains of development, whereas preschool GOMs tend to focus on language, literacy, and mathematics. Another reason is that some of the steps involved in implementing a curriculum-embedded approach contain components similar to those used for GOMs, but the reverse is not true. A measure of curriculum-embedded approaches could therefore be adapted for use with GOMs in the future, but a GOM-based measure would be difficult to rework for curriculum-embedded approaches.

Developing a measure to evaluate the ongoing use of curriculum-embedded approaches for individualization would help inform research about the importance of the different dimensions within each stage of implementation. Ultimately, this could help teachers better individualize instruction and produce better outcomes for children. In the next chapter, we describe the recommended multi-method measure.

CHAPTER III. PROPOSED MULTI-METHOD MEASURE

Our plan for measuring teachers' use of ongoing assessment is grounded in the indicators of quality identified in the conceptual framework for the curriculum-embedded approach. The plan draws on information from several sources, including the literature review, input from the expert consultant group and ACF, examples of teachers' assessment documentation obtained from the expert consultant group and other prominent researchers, and reviews of manuals for GOMs and curriculum-based assessments. Together, these sources helped us identify key constructs to measure as well as data sources for measuring them. The plan includes options for operationalizing these constructs and lays the foundation for developing and pre-testing a measure to examine teachers' use of ongoing assessment for individualization as part of this project's Optional Services Components (OSCs) 2 and 3, if these options are exercised.

We propose using a multi-method measure called the Tool for Tailored Teaching (T3). This measure will consist of a document review, video-based observations, and a one-hour teacher interview with a reflective think-aloud protocol.[12,13] We will develop scoring systems for each method. The T3 will capture the constructs in each stage of the conceptual framework while balancing the competing considerations of (1) reliability and validity; (2) burden on researchers, teachers, and classrooms; and (3) budget concerns, both for development and ongoing use.

This chapter presents our rationale for using the T3. We discuss the issues related to choosing a measurement approach, recommend data sources and methods for a multi-method measure, and highlight contextual factors that may help us interpret our findings. We also provide examples of the type of content that could appear in the measure. We then present several options for expanding the measure to inform professional development. Finally, we discuss how the T3 could complement other classroom quality observations, such as the Classroom Assessment Scoring System (CLASS) (La Paro et al. 2012).

A. The Need for a Multi-Method Approach

Experts in research design and methodology strongly recommend using a multi-method approach when measuring constructs (Brewer and Hunter 2006). Specifically, they recommend triangulated measurement (Campbell and Fiske 1959; Webb et al. 1966; Denzin 1978; Patton 2002; Ritchie and Spencer 2002; Denzin and Lincoln 2011), 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 that it intends to measure. When the different methods yield similar results, we have more confidence that the construct is being measured validly. Given the complexity of teachers' use of ongoing assessment for individualization, a multi-method approach would best enable us to cover all aspects of the process.

---

[12] Teachers will be asked to reflect and "think aloud" about how they made decisions as they conducted assessments and used the data to inform their teaching. Throughout this volume, we use the term "think-aloud" to refer to this reflective process.

[13] The terminology used to describe certain data sources and methods varies across and within research fields. To frame this discussion, we have included definitions of our key data sources and methods in Appendix A. Appendix B provides examples of prior use of these data sources.

Key Recommendation: Summary of Proposed Measure

We propose using a multi-method measure—the Tool for Tailored Teaching (T3)—to examine teachers' use of ongoing assessment. The T3 will consist of a document review, video-based observations, and a one-hour teacher interview with a reflective think-aloud protocol. Teachers will video-record their assessment activities and gather documents over a two- to three-week period, followed by a one-day visit from a researcher to conduct the document review, rate the videos, and interview the teacher.

- Document review and ratings:
  - The document review will involve gathering ongoing assessment documentation (such as a portfolio) for two children, one performing well and the other struggling, to see how teachers use ongoing assessments to individualize instruction.[14] The researchers will also review classroom-level assessment information and current lesson plans for evidence of individualization.
  - Researchers will rate the documents with rubrics, checklists, and ratings.
  - *Assessment documents*. Researchers will use different methods to code certain constructs (such as evidence of frequent data collection, objective documentation, or data organization). For example, the T3 may have a special rubric to help the researcher examine how the teacher selects an assessment target.
  - *Instruction documents*. Researchers will review and code lesson plans for evidence that the teacher's instruction was responsive to the data and for any plans for differentiation or individualization. For example, the T3 may have a checklist to identify individualized instructional adaptations in the lessons plans, such as grouping strategies, additional practice, or environmental supports.
- Classroom observations:
  - The teacher will video-record a combination of assessments and small-group instruction that includes one or both of the target children.
  - Researchers will view the video at the center after rating the documents and will analyze it using rubrics, checklists, and ratings. For example, the T3 may have a rubric to examine approaches to individualization, such as additional prompts for children, the positioning of children, additional wait time before prompting certain children, and other techniques that may not be evident in the lesson plans.
- One-hour teacher interview with reflective think-aloud protocol:[15]
  - During the interviews, the researchers will probe for additional explanations about the documents and video data as well as teachers' planning and use of adaptations, modifications, and individualized teaching strategies. Teachers will describe how they used the data to gauge children's progress or to perform other tasks (for example, to help make instructional decisions). Researchers will also ask about teachers' grouping strategies and use of classroom-level data.
  - Researchers will rate the interview data with a rating or rubric, coding the teachers' responses about their interpretation of data, any alternatives considered, their decisions about how to individualize, and the success of their efforts.

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

[15] The interview and reflective think-aloud will take place after the video-recorded observations and document review. The researchers will ask each teacher to reflect on her thinking while she collected and used the data. Throughout this report, we use "think-aloud" to refer to this reflective process.

The T3's multi-method approach can yield more complete information and thus a richer sample of teaching behaviors and decision-making. For example, documents such as assessment records or portfolios can provide more information about teachers' responses to data across a broader sample of children and domains than we could learn in a single observation. Similarly, teacher interviews can give us insight into what teachers think and say that they do, but classroom observations show us what they actually do. When a complex and multidimensional process is measured, collecting data from various sources that provide both overlapping and distinct information is critical to unpacking the components of the process.

Having teachers explain the motivation and ideas that guide their behavior and instruction is especially important, given our goal of understanding how teachers select assessment activities and make instructional decisions. We want teachers to make explicit what for them is likely an implicit application of their observations, content knowledge, knowledge of child development, and pedagogical knowledge to instruction. A multi-method approach will allow us to capture what they think, know, and do when assessing children and tailoring instruction.

Examining teachers' use of ongoing assessment data requires that we understand five factors that affect the process:

1. *Teacher knowledge and beliefs.* Teachers' familiarity with child development and learning, recommended practices, individualized instruction (in general and in each content area), ongoing assessment, and similar topics will affect how they use data to make instructional decisions. Understanding teachers' knowledge will help us analyze their decision-making. For example, to use assessment data to tailor instruction, teachers need to know how children learn certain content, including the expected trajectory of their learning and what to do if they are falling behind. A teacher's expectations for children may also provide insight into teacher beliefs about development. The interview questions will help to elicit the types and depth of knowledge that teachers call upon as they collect and use ongoing assessment data.
2. *Teacher thoughts.* Documenting what teachers are thinking as they complete assessments and individualize instruction will shed light on how they use data about each child. This information will also be helpful in understanding how teachers use their knowledge of child development, effective practices, individualization strategies, and the curriculum to make instructional decisions.
3. *Teacher practice.* Observing what teachers actually do in the classroom will help us understand how they implement curricula in general and how they collect and use ongoing assessment data more specifically. Teachers need to know how to translate their knowledge of instructional strategy into actual practice and how to adapt instruction when children are still not meeting expectations. Classroom observations are the most direct way to measure this. They also 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.
4. *Contextual factors.* To interpret our results, we may collect information on contextual factors such as program support for planning and reflection, teachers' background, and any relevant support or professional development teachers receive. For example, if a teacher is having difficulty modifying instruction based on the ongoing assessment data, it would be important to know how much and what type of professional development

the teacher has received about ongoing assessment. Context could also help us understand what factors might moderate the relationship between child outcomes and ongoing assessment for individualization. However, the benefits of having this contextual information need to be weighed against the additional time burden that collecting it would place on teachers. Exhibit III.1 presents examples of contextual information we could gather in a teacher interview.

5. *Interplay of these factors.* Understanding the interplay between teachers' knowledge, beliefs, thoughts, practices, and context will help reveal the source of strengths and challenges in ongoing assessment and individualization (such as an issue with any of these factors or the teacher's ability to reflect). This information will allow programs to provide targeted support and training to each teacher.

Our three proposed data collection methods— document reviews, classroom observations, and teacher interviews—are part of a triangulated approach to learning about teachers' knowledge and use of ongoing assessment data. Each will provide unique information, and together, they will paint a more complete picture of this complex process. For example, the teacher interview can provide an understanding of the selection and development of documents and expand on the use of the documents in the classroom. The observations, in turn, can provide information about how teachers translate their knowledge into practice. Together, the information collected can provide a more comprehensive profile of teachers' use of data to inform instruction. In the next section, we present challenges to this triangulated approach.

Exhibit III.1. Examples of Contextual and Background Information to Gather in a Teacher Interview

Contextual factors assessed in a teacher interview could include:

- The teacher's available time and the burden of collecting and using assessment data to individualize
- Feedback or support from others (such as the director, an education coordinator, or fellow teachers) or collaboration in the process (as in a team setting)
- Classroom structure (for example, whether an assistant is available to help conduct small groups and whether the schedule allows time for assessment or one-on-one individualized instruction)
- Ongoing professional development related to the collection and use of ongoing assessment data
- Teacher beliefs about the utility of assessment data to tailor instruction

Teachers' background characteristics assessed in a teacher interview could include:

- Teacher education and credentials
- Years of experience teaching (in general)
- Years of experience teaching in early childhood
- Years using current assessment tool and curriculum

B. Measurement Issues

The goal of developing a measure is 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 tell us what we want to know. In this section, we describe the components needed to develop such a measure. We then consider the feasibility of developing a measure that can be used in classrooms during a typical day—an issue that will be especially critical if the measure is used widely. We also discuss procedures for selecting classrooms and teachers as they relate to measurement development and feasibility. Our pre-test of the measure will reveal whether it can feasibly collect reliable data.

1. Balancing Validity, Reliability, and Feasibility

A valid and reliable measure for this project should include (1) multiple methods, (2) enough items or indicators to capture the breadth and depth of the constructs being measured, (3) an examination across multiple learning domains, and (4) detailed scoring rubrics for each construct. However, we need to balance these requirements with the need for feasibility, especially given the complexity of the multi-method approach and the amount of time it might take to train researchers to learn and implement the different components. After all, the utility of an approach that cannot be flexibly and easily used in a classroom is limited. Teacher and researcher burden should be the primary considerations in evaluating feasibility, along with the capacity for collecting reliable information. We discuss these considerations below, highlighting the feasibility issues for each.

Multiple methods. The more methods we use to measure a construct, the more confidence we can have in our findings (Westen and Rosenthal 2003). In this way, using a multi-method approach enhances *construct validity,* or the degree to which we are measuring what we intend to measure (Brookhart 2009). This may also enable us to use one data source to check the validity or reliability of another, providing a critical safeguard against the potential weaknesses of each. For example, document reviews alone cannot provide information about the reliability and validity of the documentation, and the validity of classroom observations may be affected if the teachers or children alter their behavior because they are being studied.[16] Used together, these methods can provide a more complete picture. In addition, using more than one method also helps us recognize the different sources of error in each form of assessment.

Feasibility of using multiple methods. Although it is valuable to collect data by many methods, the benefits of that data must be weighed against the burden imposed on teachers, classrooms, and researchers.

- The longer the measure takes to implement, the greater the burden on the teacher. Collecting the documents, video-recording classroom activities, and being interviewed all take teachers' time. This time may be during class, which may take away from instructional time or require additional personnel. Asking teachers to do these tasks after school is also a burden. During pre-testing, we will attempt to gauge feasibility by soliciting information on the time that teachers spend on these activities and their perception of the burden during debriefing. During the OSC 2 pre-testing, we will also

[16] This phenomenon is known as the Hawthorne effect (Gillespie 1991).

look for any unintended consequences due to the number of observations and teacher involvement in the video recording.

- The more methods that the researcher needs to learn and implement, the greater his cognitive burden (which can hinder reliability and validity) and time burden (which can increase costs). For example, for the documents to provide evidence of teachers' knowledge of child development, the researchers themselves must have enough knowledge of child development to rate the evidence. The cognitive burden in applying knowledge (of children and instruction) across multiple domains will be high unless researchers have extensive knowledge of these areas. This burden could also limit the pool of researchers capable of using the measure.

Multiple items within each method. Increasing the amount of information gathered via a single measurement method can increase the reliability and validity of a measure. The teacher interview, for example, could address several items: how teachers collect assessment data, how they interpret their data, and how they use the data to plan their instruction. Each item could yield meaningful information.

Feasibility of using multiple items within each method. Again, the benefits of this approach need to be weighed against the burden on teachers and researchers. For example, although reliability could be improved by including more interview questions, more questions mean a longer interview. We must therefore consider whether our project goals could be reached with fewer, more targeted questions.

Assessment across multiple domains. To understand teachers' general use of ongoing assessment data for individualized instruction, we need to measure their engagement with this process in more than one developmental domain. Ideally, we would measure the process across all five key domains in the Head Start Child Outcomes Framework: (1) cognition and general knowledge, (2) language and literacy, (3) approaches to learning, (4) social and emotional development, and (5) physical development and health (U.S. Department of Health and Human Services 2011).

Feasibility of assessing multiple domains. Using all five domains in Head Start's framework may result in a measure that is all-inclusive but infeasible. Such a measure might be too long or otherwise would gather insufficient depth of information on any given domain. Considering the budget limits, attempting to address each domain in the first phase of measurement development may also limit the possibility of refining it for any one domain. Our proposed alternative is to focus on two domains, which will allow us to carefully examine and compare teachers' use of ongoing assessment in specific areas. Understanding how teachers use data in two domains can also help inform the development of measures for other domains.

Coding scheme and holistic scoring rubric. To derive the most information from the documents, videos, and teacher interviews, we will use a coding scheme and holistic scoring rubric. Researchers will use the coding scheme to extract data from each source to inform their use of a holistic scoring rubric. Holistic rubrics allow the researcher to consider how the different indicators in a given area fit together to create quality and judge overall quality using multidimensional descriptions on a scale. So, compared with more simplistic checklists, holistic rubrics are ideal for capturing constructs when the whole is greater than the sum of the parts. Applying a holistic rubric to diverse sets of ongoing assessment data can reduce the number of items in a measure. However, holistic rubrics can be more difficult to score and require more training than one-dimensional measurement methods.

Feasibility of using a holistic scoring rubric. One challenge will be developing holistic rubrics that can triangulate the data from all three methods reliably. Although holistic scoring rubrics allow us to extract complex information from methods such as observations and interviews, they put increased cognitive demands on the researcher. The need to consider multiple constructs and methods when scoring can make it challenging for researchers to become reliable on the scoring system. To ensure strong inter-rater reliability, holistic rubrics often require more extensive researcher training compared with simpler rubrics or checklists. Training materials will need to support researchers in understanding what to look for and how to weigh different factors in scoring.

2. Accommodating a Variety of Assessment Systems

Another aspect of feasibility to consider is that the measure must accommodate a variety of assessment systems used in Head Start classrooms, including both electronic portfolios (such as COR Online, WSS Online, and Galileo) and hard-copy records of child performance and behavior. Consequently, interview questions will need to be semistructured to apply across different assessment systems, and researchers will bear a greater burden in terms of additional training about when to ask follow-up questions and how to rate. The scoring system should allow 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 expectations for progress).

3. Issues Related to Selection of Classrooms and Scheduling

Another important factor to consider is how to collect evidence of teachers' practices. We need to determine the optimal timing and frequency of data collection as well as how to select children from each teacher's class. Both factors will affect the balance of validity versus burden.

Timing and frequency. The best way to gather the information we seek is to conduct multiple observations over the course of a year (Stuhlman et al. 2010). A single observation may not reveal the full range of teachers' behaviors, especially given that the assessment and individualization process is ongoing. A series of observations, on the other hand, will help us ensure that what we observe is a valid representation of teachers' practices. The expert consultant group advised that we conduct at least three observations of each teacher during the OSC 2 pre-testing to determine what can be learned in a limited number of observations.

The timing of the observations can affect the burden on teachers and their involvement in the project. One approach is to observe a specific event at a scheduled time with the researcher present. This will yield valuable information that will allow us to triangulate the data from the documents with the observation and interview data, ultimately creating a picture of what teachers do with ongoing assessment data. However, more observations would provide more information about the process. Multiple in-person observations would be costly and might not be realistic for a measure that is brought to scale; we therefore recommend that teachers video-record their assessment and small-group instructional activities. These short 10- to 20-minute videos can be rated by the researchers during a single in-person visit. Besides being less costly, allowing the teacher to video-record gives him more control over the timing of the activities and the selection of footage

submitted to the researcher.[17] Video recording also increases the feasibility of the multi-method measure because it requires only a single visit by the researcher. We recommend attempting more than one approach to the frequency and focus of video recordings during pre-testing to determine the value of additional observations (see Chapter IV for details).

Child involvement. How to select children for the classroom observations is an important issue. One question is whether the classroom observation and document review should include the whole class or only selected children. If we focus on selected children, the procedures for choosing the children must be clearly defined, and in either case, the burden on the teacher, children, and researchers must be kept in mind.

All things considered, we believe it would be better to focus on two selected children. Purposively sampling two children is more cost-effective and would allow for more in-depth study than examining the entire classroom. Focusing on the entire class would provide ecologically valid data because teachers work with all children in the class, and it could provide evidence of how broadly teachers individualize. However, we believe that the costs of observing and reviewing documents for approximately 20 children would outweigh the benefits. Focusing instead on a few children would yield thorough data on those few, but it may not allow us to test the measure across the full range of teacher and child behavior. To address this, we recommend selecting two children—one high-performing and one struggling—and investigating how the teacher uses ongoing assessment data for both.

C. Recommended Multi-Method Approach

The T3's multi-method measurement protocol will include (1) a review of documents, (2) video-based classroom observations, and (3) a teacher interview. Teachers will create recordings over a two- to three-week period, after which the researcher will conduct a one-day visit to review the documents, watch the video, and interview the teacher. As shown in Figures III.1, III.2, and III.3, this multi-method approach allows for triangulation across data sources. That is, information collected can be used to confirm data collected from another source or to fill in gaps in another data source. The T3 will have scoring rubrics that we will develop and refine through pre-testing (see Chapter IV). We anticipate the T3 will consist of 12 or fewer rubrics across the data sources, accompanied by checklists or ratings. Overall, our goal will be to develop the T3 such that it can be used flexibly and easily by researchers, education coordinators, master teachers, coaches, and teachers to yield high-quality data. We discuss each T3 method in detail below.[18]

[17] Although 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 what we can learn about actual teacher practices, the footage can still capture teachers' knowledge of assessment because, to select the best footage, they must understand what constitutes good assessment practice.

[18] Please note that the T3 will measure how well a teacher conducted ongoing assessments for individualization regardless of the assessment tool used. In other words, the T3 will rate the quality of a teacher's use of the tool, rather than the quality of the tool itself.

1. Document Review

Researchers will review a number of documents—such as student portfolios, assessment records, and lesson plans—to measure the intentionality, focus, completeness, and objectivity of the teacher's data collection and instructional planning (Exhibit III.2). This review can also address constructs from each of the four stages in the conceptual framework. Teachers will provide us with documentation for two children, one performing well and the other struggling. We will give each lead teacher clear guidelines explaining the types of documents we would like to see.

Figure III.1. Multi-Method Measure Model: Document Review

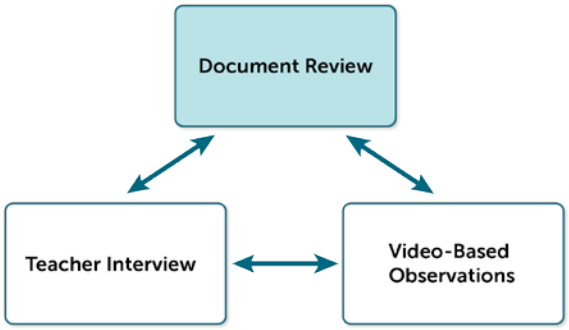


Exhibit III.2. Examples of Documentation

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

Below, we describe the steps researchers will use to collect documents. This process can be adjusted based on what we learn from the pre-test (see Appendix C for details).

Step 1. Selecting children. We will begin by pre-testing a method for selecting two target children (one performing well and the other struggling) for whom the documents will be collected. First, we will seek consent from the parents of all children in the classroom using a consent form approved by the Institutional Review Board (IRB).[19] We will then ask the teacher to complete the Oral Language & Comprehension measure, or OL&C, for the children with parental consent (see Appendix D) (Bradfield and McConnell 2013). This measure has only eight items per child and takes about 10 to 15 minutes to complete for an entire class.[20] The OL&C is specific to the domains of language and literacy and does not require training. Although the teacher's results will be subjective and could be influenced by many factors, researchers have successfully used the OL&C in combination

[19] We will seek parents' consent for the teacher to complete the OL&C measure (Bradfield and McConnell 2013), share documentation about the target children, and participate in video-based observations. We will also seek permission to use these videos for team discussion, item refinement, and researcher training.

[20] A potential issue is the teacher burden associated with asking teachers to spend 10 to 15 minutes rating the children. However, this approach may also provide informative data on how teachers view students and the teachers' decision-making process.

with IGDIs to identify children in need of instructional intervention.[21] We will use the teacher's results to rank the children and will randomly select one child who is performing well and one child who is struggling.[22] After selection, we will call the lead teacher to inform her of our selections and discuss next steps.

Step 2. *Assembling documents.* As part of our recruitment efforts, we will send letters to the lead teachers that include a description of the documents we would like them to collect before we visit. The letters will specify that we would like to see any documentation related to the teachers' assessments, plans, and lessons for the identified children. We will also provide a list of sample documents.

Step 3. *Analyzing the data from the document review.* The review will take a researcher about one hour to complete during the one-day visit to the program. The researcher will use an instrument that we will develop to abstract information from the documents. This instrument will record the type and frequency of assessments, ways in which data have been used to guide instruction, and any evidence of individualization in lesson plans. The instrument will allow researchers to determine if the documentation for each child provides evidence of the child's unique strengths and weaknesses as well as common areas of progress across children. Researchers will check for information across time about specific areas of development, rather than a random collection of skills and behaviors that does not focus on progress.

In general, we recommend using 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.[23] For example, even if information is gathered frequently, regularity alone is not a sign of good assessment unless the information is about areas that are expected to change, can be observed, are documented reliably, and lead to valid interpretations. A holistic rubric can tease out these qualitative aspects of the assessments. Exhibit III.3 includes questions to consider in developing rubrics and ratings for the document review (presented for each of the four stages in the curriculum-embedded conceptual framework).

[21] Judith Carta, personal communication by email, July 10, 2013.

[22] We will also select an alternate for each target child in case one of the children cannot participate (for example, if the child is absent during the video recordings) or if a parent refuses permission. This measure has not been previously used by itself to identify children who are struggling and those who are performing well, and we will consider its use for this purpose during the pre-test.

[23] Researchers will also rate documents using ratings or checklists when the presence or frequency of certain behaviors is important (for instance, whether the teacher is providing regular progress reports to families).

Exhibit III.3. Examples of Questions to Be Addressed by the Document Coding Scheme, by Stage of the Conceptual Framework

STAGE 1—SELECTING TARGET AND ASSESSMENT METHOD FOR ASSESSING CHILD PROGRESS
1. What type of goal for child learning is being targeted—a content goal related to what a child knows (such as naming numbers, letters, and shapes) or a process goal related to how a child acquires knowledge (such as identifying the main ideas of a story)? 2. Are the tasks for activities used for ongoing assessment connected to the curriculum, the child's interest, or a specific assessment system? 3. Does the teacher use multiple contexts over time to assess a single skill, or does she use a standard probe across time to chart change? 4. How often does the teacher collect information on the child's knowledge and skills?
STAGE 2—IMPLEMENTING ONGOING ASSESSMENT OF CHILD PROGRESS
1. What is the method and content of documentation? • What evidence is collected, how, and how efficiently? • How well-aligned are the data collection and documentation with the goal of the ongoing assessments? • To what extent are specific child behaviors documented? 2. Is the documentation connected to a curriculum? 3. Are the observations documented with descriptions of random activities without a clear goal ("6/22 Judy played with Johnny in the sand"), or is there an identified area of learning with sufficient information across time to support inferences about children's progress ("6/22 Judy built a castle with Johnny in the sand using language to plan the castle features ['We need more sand here,' 'Let's put a window here']")? 4. What evidence do we find in the documentation regarding the teacher's knowledge of child development or of the knowledge of individual children? 5. Does the teacher focus observations on specific objectives and use information from one observation to inform the next?
STAGE 3—INTERPRETING DATA AND MAKING INSTRUCTIONAL DECISIONS
1. Is the evidence organized in ways that support charting progress? 2. What evidence does the teacher use to justify decisions or interpretations? 3. What other sources of information does the teacher use to interpret findings (for example, information provided by the special education teacher, parents, etc.)?
STAGE 4—APPLYING INSTRUCTIONAL DECISIONS AND INDIVIDUALIZING
1. How does the teacher use data on child performance and progress to change or adapt instruction for a specific child? 2. Is there evidence of planning based on previous documentation? 3. What adaptations and modifications are evident in the activities and instruction? • How does the teacher group the children? • What environmental supports are evident? • What peer strategies are evident? • What instructional interactions does the teacher use to support each child's needs?

The document review will focus on two types of documents: assessment documents (such as plans for assessments and assessment results) and instructional documents (such as lesson plans). For assessment documents, the constructs and method of coding will be based on the various aspects of documentation (for example, evidence of frequent data collection, objective documentation, and data organization). Researchers will use the rubric shown in Exhibit III.4 to examine teachers' selection of an assessment target (stage 1 in the conceptual framework). This rubric is designed to help quantify teachers' use of child data along a seven-point scale, delineating the specific behaviors a teacher must demonstrate to achieve a particular score. We will pre-test this rubric, along with others that will be developed in OSC 2, to ensure that the overall categories and the specific target criteria provide the information needed to understand teachers' use of ongoing assessment data.

Exhibit III.4. Rubric to Examine the Selection of Assessment Targets

STAGE 1—SELECTING MEANINGFUL ASSESSMENT TARGETS			
Assessment targets are linked to meaningful outcomes (that is, teachers assess skill, knowledge, or behavior that a child needs in order to be successful now or in the future). Targeted behaviors are developmentally appropriate (for example, looking at combinations of five with objects rather than asking children to solve written equations). To meet the requirements of Head Start and other early childhood education policies, 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 ones in which children of this age typically make progress within the program year (that is, change is expected in the current year). The targets address areas that are taught in the classroom curriculum. The targets are defined and measured based on observable behaviors. The targets can be assessed universally (that is, for all children), or evidence of individualization is clear, with response modes available for the children to demonstrate progress.			
1	3	5	7
Target is defined but not clearly linked to a structured curriculum or to meaningful outcomes. OR Target is not developmentally appropriate.	At least one target is defined and linked to a structured curriculum. At least one target is defined for one of Head Start's five key school-readiness domains. Target is measured using observable behaviors. Target may not be in an area in which children typically make progress within the program year (that is, change cannot be measured over time).	A few targets are defined, linked to a structured curriculum, and can be changed with instruction or intervention. At least one target is defined for each of Head Start's five key school-readiness domains. Targets are measured using observable behaviors. Targets are in an area in which children typically make progress within the program year (that is, change can be measured over time).	Multiple targets are defined, linked to a structured curriculum, and can be changed with instruction or intervention. Multiple targets are defined for several of Head Start's five key school-readiness domains. Targets are measured using observable behaviors, AND teacher considers whether targets generalize across settings. Targets indicate amount of progress expected within time intervals (that is, change can be measured over time).

Notes: Definitions of terms and examples will be provided in a researchers' training manual to facilitate reliable completion of each rubric. Ratings of 2, 4, and 6 should be assigned to teachers who exceed the criteria for the next-lowest rating but do not yet meet all criteria for the next-highest rating. For example, if all the criteria for a 1 were met but not all the criteria for a 3, the rating would be a 2.

For the instructional documents, researchers will review and code teachers' lesson plans for evidence that their instruction is responsive to the data and any associated plans for individualization. Exhibit III.5 presents a checklist to examine lesson plans for evidence of individualization (related to stage 4 in the conceptual framework). We will pre-test the checklist to ensure that it captures all essential information.

Exhibit III.5. Checklist to Examine Lesson Plans for Evidence of Applying Instructional Decisions and Individualizing

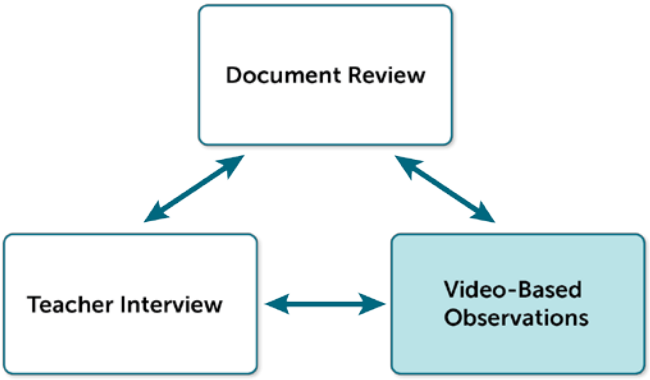
STAGE 4—APPLYING INSTRUCTIONAL DECISIONS AND INDIVIDUALIZING	
Lesson plan demonstrates evidence of varied approaches to individualization that are responsive to the data **MARK ALL THAT APPLY** *If any items are marked, probe during teacher interview*	
• Adaptations or modifications designed to meet specific child needs	☐
• Prompting or questioning strategies	☐
• Additional practice	☐
• Grouping strategies	☐
• Peer supports	☐
• Adult supports	☐
• Environmental supports	☐
• Other *(please describe)*	☐

Note: To help researchers reliably complete each item-level rubric, the researcher training manual will include definitions of terms and examples.

2. Video-Based Classroom Observations

Video-recorded observations will give us a critical piece of information about how teachers conduct assessments and use the data from the document review to inform their instructional practice. (To augment what the research team sees in the video, the interview will include a reflective think-aloud component to provide more insight on the teacher's thoughts and actions.) The videos could be used to provide information about stages 2 and 4 of the conceptual framework, which focus on teacher use of assessments. Our proposed steps for conducting these observations, described below, can be adjusted based on the information learned from pre-testing.

Figure III.2. Multi-Method Measure Model: Video Observation



Step 1. Learning to use the equipment. Teachers will be shipped a tablet and accessories (including a charger and small tripod) and will receive video-based and written instructions for recording. We will encourage them to experiment with the equipment before recording a formal video. The team will also have a call with the teachers to answer their questions and will be available to offer technical assistance as needed.

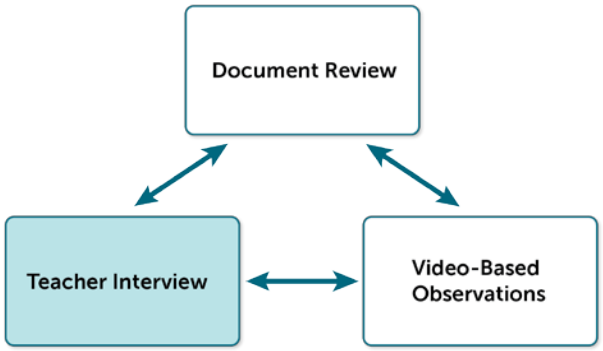
Step 2. Collecting the videos. In each classroom, the lead teacher will record a selection of activities involving individual ongoing assessment and small-group instruction with the target children. We recommend documenting instruction over two to three weeks. The recordings will be completed before researchers visit the classroom.[24] We will pre-test two procedures that vary the focus and frequency of the video recording (see Chapter IV for details). Teachers will use the assessment tool that they normally use. For the small-group instruction, we will ask that language or literacy be the focus, but otherwise we will not put parameters on the activities, instead asking teachers to follow their typical classroom practice.

Step 3. Coding the data from the videos. We will develop rubrics for the researchers to use in coding the video-recorded observations. The rubrics will be used to examine the degree to which teachers were able to collect data and individualize instruction. Each researcher will view and code the recordings during the one-day visit to the classroom (after the document review and before the teacher interview). As with all rubrics, the items will be refined based on pre-testing. The tablets will be collected at the end of the visit, and the videos will be available to the assessment team for further discussion.

3. Teacher Interview

Researchers will interview the lead teacher at the end of the one-day visit to the classroom.[25] This teacher interview will include a reflective think-aloud protocol and will provide additional information about the documents and the video-recorded observations to help the assessor understand the teacher's thinking and decision-making while collecting data and using the data to inform practice. The researcher will use a standard set of questions about the teacher's use of ongoing assessment data to inform instruction. During the reflective think-aloud portion of the interview, the researcher will ask the teacher to describe her thoughts and actions as she completed the video-recorded tasks. This will shed light on the teacher's decision-making processes.

Figure III.3. Multi-Method Measure Model: Teacher Interview



[24] We will ask teachers to check that the video was actually recorded (that is, the footage is not blank).

[25] During pre-testing, we will ask selected teachers to comment on the length of the interview during a post-interview debriefing. We will schedule the researcher's visit at the teacher's convenience, preferably on a day when an assistant can cover for the teacher being interviewed. Conducting the interview during children's naptime may also be feasible, and the expert panel encouraged conducting the interviews in the classroom when possible. The researcher could also visit the classroom on a teacher in-service day or during the evening.

The one-hour teacher interview will probe for additional information about assessment data obtained as part of the document review and observations, as well as about teachers' planning and use of adaptations, modifications, and individualized teaching strategies. Teachers will describe how they determined the goal of an assessment document, collected and interpreted the data, and used the data to gauge children's progress or for other purposes (for example, to communicate with families). The interview questions could also address other types of evidence or documentation teachers use and how and when they use them. Proposed steps in the interview procedures are described below and can be adjusted based on the information learned from pre-testing.

Step 1: Interviewing the teacher. During the interview, we will ask teachers to think aloud about the instructional decisions they made during the video-recorded activities. We will develop a set of standard, semi-structured, open-ended questions and prompts to use during this think-aloud period. We will also ask questions (with follow-up prompts) about teachers' data collection strategies, their understanding and use of their assessment tool, what they learned about the children in their class, and their use of data to tailor their instruction for individuals and small groups. For example, we will ask how the teacher responded when a strategy proves to be ineffective.

Additional questions will probe teachers' knowledge and decision-making, supplementing information from the document review and video-based observations. These questions will be organized by the researcher before the interview starts. Researchers will ask:

- What the documents obtained as part of the document review reveal about a child's abilities and any steps the teacher did or planned to do to support the child's learning based on that understanding. Researchers will use additional probes, taking care to avoid leading teachers and skewing the results.
- For explicit examples of how teachers used ongoing assessment data, specifically listening for examples of individualizing instruction and communicating with families.
- Why teachers collected the particular information they had (listening for intentionality and how it relates to the curriculum and to each child's needs); how they interpreted the data; and what actions they took, if any, in response to what they observed.
- Whether teachers shared their information with parents or involved them in collecting any data. If they do share the data, researchers will ask for specific examples of how they did 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).

Step 2: Coding the interview responses. Researchers will record teachers' responses to the think-aloud and interview questions.[26] They will later code and score these responses, primarily using holistic rubrics. The teacher interview and think-aloud will take about an hour to complete, plus an hour for the researcher to score. Before leaving the center, researchers will code interview responses and

[26] During the pre-testing phase, with the teacher's permission, we will record the interviews for later analysis by the team.

teachers' notes and comments on their interpretation of data, any alternatives considered, their decisions about how to individualize, and the success of their efforts.

The coding protocol for the think-aloud and semi-structured interview questions could be framed around the following questions:

- How does the teacher schedule observations or the use of a specific ongoing assessment tool?
- How does the teacher select assessment strategies for different children and organize instruction?
- What does the teacher look for when observing children? How focused are the observations?
- How does the teacher interpret children's errors and misconceptions? What does the teacher do about the errors and misconceptions?
- Does the assessment reflect the teacher's attempt to gauge what the child knows and doesn't know?
- How does the teacher use the information that she documents? How does she organize it for interpretation, and how does she plan what to do next with the child?
- Does the instruction include attempts to teach the same concept in different ways?
- Does the instruction show that the teacher is building on the child's strengths to scaffold the child's learning?
- Does the teacher use any technology to support ongoing assessment?
- How does the teacher share information with families and involve families in collecting information?

Exhibit III.6 presents a rubric to rate teacher interview responses and instructional documents after probing for examples of individualization (related to stage 4 in the conceptual framework).

Exhibit III.6. Rubric to Examine Teacher Interview Responses and Lesson Plans for Application of Instructional Decisions and Individualization

STAGE 4—APPLYING INSTRUCTIONAL DECISIONS AND INDIVIDUALIZING						
Using a variety of approaches to individualization that are responsive to the data. The approach is responsive to the data, the interpretation of the data, and the child's progress. The teacher collects further evidence to examine whether a given instructional approach is valid for meeting the targeted child's instructional need. 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, giving a child more practice in a particular area, 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 also provide peer, adult, and environmental supports. Whenever possible, the teacher incorporates and builds on children's strengths and interests when individualizing.						
1		3		5		7
Provides standard instruction for all children, with no evidence of individualization or differentiation.		Increases opportunities for children with weaknesses to practice their skills. Adapts instruction for at least one child.		Increases or varies opportunities for children with weaknesses to practice their skills. Uses some varied instructional strategies. Provides support for emerging skills identified in the data, recognizing when children need a challenge.		Increases or varies opportunities for children with weaknesses to practice their skills. Uses varied instructional strategies to build on strengths and mitigate weaknesses. Incorporates child interests and experiences into instruction. Collects data on effects of individualization strategies to determine influence on child's growth. Reflects on the success of instructional approaches. Reaches out to external resources as needed. Classroom instructional team has a shared knowledge about each child's goals and the instructional strategies.

Notes: Definitions of terms and examples will be provided in a researchers' training manual to facilitate reliable completion of each rubric. Ratings of 2, 4, and 6 should be assigned to teachers who exceed the criteria for the next-lowest rating but do not yet meet all criteria for the next-highest rating. For example, if all the criteria for a 1 were met but not all the criteria for a 3, the rating would be a 2.

D. Initially Targeting Two Domains

All five domains in the Head Start Child Outcomes Framework are important for understanding child progress and are therefore relevant to our study. But there are several constraints that prohibit exploring all of these domains in the initial phase of measurement development (OSC 2). Perhaps most prominent is the burden on teachers, classrooms, and researchers when data are collected in many domains.

Instead of imposing this burden, the expert panel suggested pre-testing 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). Head Start has also invested considerable resources in training teachers to teach language and literacy; given this investment, it will be 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), as a child with social and emotional problems may also have difficulty learning. In addition, early childhood teachers are more likely to deliver high-quality instruction in language and literacy than in mathematics, and they typically stress literacy and social and emotional development (National Research Council of the National Academies 2009). Thus, the teachers we are targeting will likely provide enough instruction in these areas for us to observe variability. Finally, limiting the number of domains will allow us to thoroughly refine the measurement of those domains while staying within our budget.

Despite limiting the focus to two domains, we will include items in the T3 to examine whether teachers are drawing on all domains to interpret child data and individualize instruction (for example, whether a teacher is employing strategies that encourage persistence—part of the approaches-to-learning domain—during a literacy activity). In the future, the T3 could be expanded to cover assessment of additional domains. This will help reinforce the message of a "whole child" approach and avoid the impression that language, literacy, and social and emotional development are the only important domains.

E. Summary of the Recommended Multi-Method Measure

Using multiple methods can paint a more accurate picture of a multifaceted process. The T3's approach, which includes documents reviews, video-based observations, and teacher interviews, allows us to capture the complexities of using ongoing assessment data to inform instruction.

The documents, for example, can provide data about how teachers plan for instruction. They may also shed light on teachers' understanding of how the content of activities as well as their structure (such as the use of small groups) can be used to tailor instruction. However, documents alone without teacher feedback will not provide the rich information needed to understand how the documents are being used to inform instruction. In a similar way, the video-recorded observations will provide insight into teachers' actual instructional practice. However, incorrect inferences could be drawn from the videos if we do not include teacher reflection and input about the observations. The teacher interview will add this critical insight. Ultimately, each source will contribute both unique data and overlapping information that together describe how teachers collect ongoing assessment data and use that data to individualize instruction.

F. Optional Approaches to Inform Professional Development[27]

Information gleaned from the T3 can be used to inform an individualized professional development plan for each teacher. This is because the T3's main function—identifying and quantifying how teachers are using ongoing assessment to inform their instruction—also provides a valuable window into teachers' strengths and weaknesses in their knowledge and practice. This information can be used to develop a plan targeting what teachers need to learn to better individualize their instruction for each child.

Although this work falls outside the scope of the current contract, ACF may consider the options below for future work or other research. These options may be costly and time consuming but could also greatly inform professional development.

1. Pedagogical Content Knowledge (PCK) with Scenario Probes

PCK questions capture the intersection of a teacher's knowledge of child development, assessment, and instruction. Besides revealing what teachers know about ongoing assessment, PCK questions could incorporate scenarios with probes about decision-making processes. These probes can be designed to isolate what teachers know from the contextual circumstances that may affect what they do. The scenario questions may 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 also be used to examine constructs from each of the four stages in the conceptual framework because teachers' knowledge is always a relevant factor. Exhibit III.7 provides an example of a PCK multiple-choice question.

Exhibit III.7. Example of a PCK Assessment Item

STAGE 1—SELECTING MEANINGFUL ASSESSMENT TARGETS
Select the best (valid, aligned with curriculum, developmentally and culturally appropriate, able to be affected by instruction and intervention, meaningful, feasible) early literacy target for preschool children (4-year-olds): a. Writes own first and last name b. Writes five or more letters ("I m hpe") or words ("I am happy") to communicate meaning c. Writes five or more different letters from dictation d. Scribbles include letter-like figures e. Copies alphabet *Correct answer is B or C, depending upon the curriculum.*

2. 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 describe what hypotheses they can make about the child's development and about how she would tailor their instruction for that child. Standard tasks could require a teacher to

[27] At this time, we are not recommending these optional approaches in the measurement development activities conducted as part of OSC 2. We look forward to discussing possibilities for these optional approaches with ACF in the future.

explain how she would group children for differentiated instruction. Standard tasks could be used to examine constructs from the four stages in the conceptual framework.

3. Advantages and Challenges of Optional Approaches

Both PCKs and standard tasks have a number of advantages. They both tap into what teachers know and think about using ongoing assessment for individualization. And so if teachers are not implementing ongoing assessment (as indicated by their scores on the T3 rubrics), then PCKs and standard tasks could help clarify the aspects of ongoing assessment for which the teacher needs additional training.

Another advantage is that PCKs (with scenario probes) and standard tasks are not limited to the children in a teacher's classroom. Since a given teacher's classroom may not include children at varying levels of performance, document reviews and observations may not enable us to compare a teacher's abilities to work with different kinds of children relative to other teachers. PCKs 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 the teacher implements that knowledge.

The hypothetical nature of these approaches creates three major drawbacks. 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. Besides creating multiple items for each dimension, a PCK test or standard task may require creating multiple equivalent forms so that teachers cannot share answers and are not given the same question or task repeatedly (which could skew the results). Third, although these methods may be a useful complement to the T3, they are not sufficient to understand what teachers actually *do*. They are therefore not included in the T3.

G. Complementing Other Classroom Observation Measures

The National Center on Quality Teaching and Learning's Framework for Effective Everyday Practice depicts Head Start's approach to school-readiness as a "house." The house has a foundation of engaging interactions and environments, a first pillar of research-based curricula and teaching practices, a second pillar of ongoing assessment of child progress, and a roof representing highly individualized teaching and learning. Observation tools such as the CLASS (Pianta et al. 2006), Early Language and Literacy Classroom Observation (Smith et al. 2008), and Early Childhood Environment Rating Scale-Revised (Harms et al. 1998) are used to evaluate the "foundation," whereas the T3 will measure the "pillars" and "roof." Used together, these tools will encompass a comprehensive approach to measuring the educational components that build school-readiness.

There is some overlap between the T3 and other measures of teachers' instructional interactions with children, such as the CLASS (especially the instructional support domain). Both measures can be used to inform improvements in the quality of instruction. However, the CLASS focuses on quality instructional interactions at the classroom level and does not address assessment and individualization. The T3, in contrast, will focus on using ongoing assessment data to individualize instruction.

CHAPTER IV. CONCLUSIONS AND NEXT STEPS

Despite the importance of using ongoing assessment data to inform instructional practice and individualize teaching, information on how early education teachers actually collect and use these data is sparse. This project can make a vital contribution to the early childhood field—especially Head Start—by providing insight into the use of ongoing assessment to deliver high-quality, individualized instruction.

The T3, in particular, could provide information on how teachers use data derived from curriculum-embedded approaches to understand children's development and to plan and individualize instruction. The final, validated version of the T3 could be used for an array of purposes by researchers, sponsoring agencies, and individual program administrators, teachers, mentors, and coaches as well as by networks of programs. Researchers can use the T3 to help staff at Head Start and in the early childhood field understand the process of using ongoing assessment for individualization. The experts we spoke with repeatedly noted that we lack even basic information about how early childhood teachers use ongoing assessment data to individualize instruction. The T3 therefore offers an unprecedented opportunity to improve educators' understanding of a process that is valued and even mandated but previously has not been measured.

Beyond basic research, the T3 might be used for a host of other purposes, such as:

- Identifying the need for technical assistance or professional development. For example, if a teacher is able to identify high-quality practices but is not implementing those practices correctly, this measure may indicate the need for coaching on those practices or education about their benefits.
- Understanding whether and how programs are meeting Head Start's requirements to conduct ongoing assessments in order to support children's learning and development (see 45 CFR 1304.20[b], 1304.20[d], and 1304.20[e]).
- Understanding the link between a teacher's use of best practices and child outcomes. For example, program managers could investigate whether high scores on the T3 are related to greater child competencies at the completion of a program year.

It is important to note that the T3 is intended to be an assessment and not an intervention. However, by articulating the relevant constructs and providing examples of a range of best practices in assessment, interpreting data, and individualizing instruction, the T3 and associated documentation could also be educational. Ultimately, the T3 could help teachers better understand the ongoing assessment and individualization process and thereby develop better practices in the classroom.

Next Steps. In the second phase of this project, we will develop and pre-test the T3. This will include the following six activities:[28]

1. Develop recruiting materials for the pre-test sample (such as lists of potential settings, advance letters, scripts, and instructions for teachers).
2. Develop an IRB package for the agreed-upon pre-test design and submit it to IRB in January 2014.
3. Draft initial items, protocols, and procedures to measure each of the four stages in the curriculum-embedded conceptual framework. We will submit these drafts to ACF within four months of ACF's exercising OSC 2 (on or around February 1, 2014) and submit the final drafts for pre-testing one month after receiving ACF's comments.
4. Test and refine the T3 via three rounds of iterative pre-tests between April and September 2014. We will conduct the pre-tests in purposively selected preschool programs with a moderate to high range of implementation of ongoing assessment and individualization.
5. Consult with experts throughout the measure development process as needed, and hold at least one conference call with the expert panel to discuss the T3. This call will be conducted in summer 2014, after the second round of pre-test data collection is complete.
6. Prepare a final report for ACF (after completion of the pre-test in fall of 2014).

The proposed pre-testing in OSC 2 will consist of three rounds of data collection. As shown in Table IV.1, the first round will involve 2 classrooms in each of 2 centers, the second round will involve 2 classrooms in 2 additional centers, and the final round will involve 2 classrooms in a single center. Ultimately, the pre-test will include visiting five centers and 10 classrooms. The centers will be chosen purposively from centers in New Jersey and/or Maryland based on the center's use of a curriculum-embedded ongoing assessment system. Teachers will receive an incentive for participating.

[28] All dates are approximate, based on ACF exercising OSC 2 on October 1, 2013.

Table IV.1. Rounds of Data Collection for the Pre-test

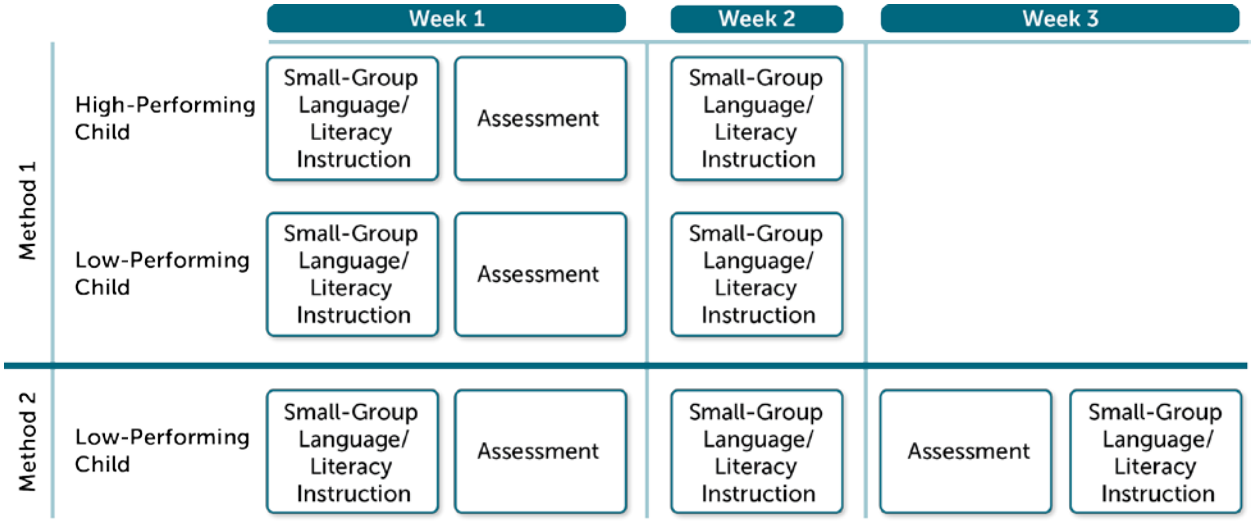
Center	Classroom	Video Method
Data Collection Round 1—Late April 2014		
Center A	Classroom 1	Method 1
	Classroom 2	Method 1
Center B	Classroom 3	Method 2
	Classroom 4	Method 2
Data Collection Round 2—May 2014		
Center C	Classroom 5	Method 1
	Classroom 6	Method 1
Center D	Classroom 7	Method 2
	Classroom 8	Method 2
Data Collection Round 3—September 2014		
Center E	Classroom 9	Recommended method
	Classroom 10	Recommended method

During this iterative process, we will test two approaches to the focus and frequency of the video recording (Figure IV.1):

- **Method 1:** The classroom teacher is asked to video-record the following sequence for one high-performing and one low-performing child, collecting six data points in two weeks:
  - Week 1: Small-group language/literacy instruction and an assessment
  - Week 2: Small-group language/literacy instruction
- **Method 2:** The classroom teacher is asked to video-record the following sequence for one low-performing child,[29] collecting five data points in three weeks:
  - Week 1: Small-group language/literacy instruction and an assessment
  - Week 2: Small-group language/literacy instruction
  - Week 3: An assessment and small-group language/literacy instruction

[29] Compared with high-performing children, instructional changes made for low-performing children should be more obvious to the observer. It should also be easier to see the progress associated with the instructional changes across this brief time period with low-performing children. Please note that teachers often attend more to children who are struggling, so the results from method 2 may not generalize to typically developing children. Information about teacher performance with typically developing children would require further study.

Figure IV.1. Two Alternative Methods for the Focus and Frequency of Pre-test Video Recordings



As shown in Figure IV.1, we will assign method 1 or 2 to each center for data collection rounds 1 and 2. In round 3, we will only use the method that appears most promising for yielding sufficient information while minimizing the burden on the teacher.

The basic steps for recruiting classrooms and conducting the assessments are listed in Appendix C. During the pre-test, we will send two team members to the classrooms. They will be professional staff with experience using observation instruments in early childhood classrooms. Although they will score the instruments independently, they will collaborate in planning the questions for the teacher interview, both be present during the teacher interview, debrief after the classroom visit, discuss scoring discrepancies, and come to consensus on item-level scores.

After each round of data collection, the full team will debrief and consider changes to the items, protocols, and procedures (including evaluating the two approaches to video-recording frequency). We will submit our recommended changes to ACF for review and incorporate any feedback. This process will be repeated for each round of pre-testing to iteratively develop and refine the T3.

With ACF's approval, we will also share the T3 with the expert consultants for their review when deemed appropriate. For example, we could hold a conference call with the expert panel between the second and third rounds of data collection. At that time, we could present a summary of the pre-test procedures; describe the duration of each measure (for example, the average length of the teacher interview); and summarize the lessons learned from the first two rounds of data collection, along with the implications for the items and procedures. We would then incorporate the expert panel's feedback during the final round of pre-test data collection.

At the conclusion of the pre-test, we will submit a draft of our final report for ACF review. This report will include the following:

- The purpose of the T3
- A description of the iterative measure development process, from the literature review to the three rounds of data collection

- Details about the T3, including descriptions of procedures and estimates of burden on teachers and programs
- A proposed training plan, including lessons learned from the pre-test
- A plan for piloting and psychometric testing of the T3
- The anticipated cost of using the T3 in a research study
- The anticipated challenges related to the reliability and validity of the T3

Developing the T3 could be an important step toward building the knowledge base on how early childhood teachers use ongoing assessment. This multi-method measure would give us a window into teachers' use of curriculum-embedded approaches to guide instruction and individualization, which in turn would help inform research about the importance of the different constructs within each stage of implementation. The findings from this study could ultimately help early childhood settings more effectively meet children's needs and produce better outcomes for children.

This page is left intentionally blank

REFERENCES

Advisory Committee on Head Start Research and Evaluation. "Final Report Submitted to the Secretary of the U.S. Department of Health and Human Services." Advisory Committee on Head Start Research and Evaluation, 2012.

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, and Evaluation, 2014.

Al Otaiba, S., C. Connor, J. 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.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing. "Standards for Educational and Psychological Testing." American Educational Research Association, 1999.

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

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

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

Bagnato, S., J. Salaway, and H. Suen. "Pre-K Counts in Pennsylvania for Youngsters' Early School Success: Authentic Outcomes for an Innovative Prevention and Promotion Initiative. Research Results of SPECS for Pre-K Counts: An Independent Authentic Program Evaluation Research Initiative (2005-2009). 2009 Final Research Report." 2009. Available at http://www.heinz.org/UserFiles/Library/SPECS%20for%20PKC%202009%20Final%20Research%20Report%20113009.pdf. Accessed March 4, 2013.

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.

Bennett, Linda. "Authentic Instruction and Assessment." *Social Studies and the Young Learner*, vol. 23, no. 3, 2011, pp. 86–92.

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

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

Black, P., and D. Wiliam. "Inside the Black Box: Raising Standards Through Classroom Assessment." *Phi Delta Kappan,* vol. 80, no. 2, 1998a, pp. 139–148.

Black, P., and D. Wiliam. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy & Practice,* vol. 5, no. 1, 1998b, pp. 7–74.

Bodrova, Elena, and Deborah Leong. "Of the Mind: A Case Study of Implementing the Vygotskian Approach in American Early Childhood and Primary Classrooms." UNESCO International Bureau of Education, 2001.

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

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

Brookhart, S. "Accountability Policies and Measures: What We Know and What We Need." Washington, DC: National Education Association Research Department, 2009. Available at http://www.nea.org/assets/docs/HE/AccountabilityPoliciesandMeasures.pdf. Accessed July 15, 2013.

Buysse, V., and E. S. Peisner-Feinberg. *Handbook of Response to Intervention in Early Childhood.* Baltimore, MD: Paul H. Brookes Publishing Co., Inc., 2013.

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.

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.

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

Carta, Judith, Charles Greenwood, Gayle Luze, Gabriel Cline, and Susan Kuntz. "Developing a General Outcome Measure of Growth in Social Skills for Infants and Toddlers." *Journal of Early Intervention*, vol. 26, no. 2, 2004, pp. 91–114.

Carta, J., C. Greenwood, D. Walker, and J. Buzhardt (eds.). *Individual Growth and Developmental Indicators: Tools for Monitoring Progress and Measuring Growth in Young Children*. Baltimore: Paul H. Brookes, 2010.

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.

Deno, S. "Whether Thou Goest . . . Perspectives on Progress Monitoring." In *Issues in Educating Students with Disabilities,* edited by J. Lloyd, E. Kameenui, and D. Chard. Mahwah, NJ: Lawrence Erlbaum, 1997, pp. 77–99.

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

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

Early Learning Standards Task Force and Kindergarten Assessment Work Group Pennsylvania BUILD Initiative. "Early Childhood Assessment for Children from Birth to Age 8 (Grade 3)." Harrisburg, PA: Pennsylvania Departments of Education and Public Welfare, 2005.

Foegen, A. "Progress Monitoring in Middle School Mathematics." *Remedial and Special Education,* vol. 29, no. 4, 2008, pp. 195–207.

Foegen, A., C. Jiban, and S. Deno. "Progress Monitoring Measures in Mathematics: A Review of the Literature." *Journal of Special Education,* vol. 41, no. 2, 2007, pp. 121–139.

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

Fuchs, L., D. Fuchs, and N. Bishop. "Instructional Adaptation for Students at Risk for Academic Failure." *Journal of Educational Research,* vol. 86, 1992, pp. 70–84.

Fuchs, L., D. Fuchs, C. Hamlett, and R. Allinder. "Effects of Expert System Advice Within Curriculum-Based Measurement on Teacher Planning and Student Achievement in Spelling." School Psychology Review, vol. 20, 1991a, pp. 49–66.

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

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.

Gillespie, Richard. *Manufacturing Knowledge: A History of the Hawthorne Experiments*. Cambridge, UK: Cambridge University Press, 1991.

Goertz, M., L. Nabors Oláh, and M. Riggan. "From Testing to Teaching: The Use of Interim Assessments in Classroom Instruction." CPRE Research Report # RR-65. Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education (CPRE), 2009.

Good, III, R., Deborah C. Simmons, and Edward J. Kame'enui. "The Importance and Decision-Making Utility of a Continuum of Fluency-Based Indicators of Foundational Reading Skills for Third-Grade High-Stakes Outcomes." *Scientific Studies of Reading,* vol. 5, no. 3, 2001, pp. 257–288.

Greenwood, C., J. Carta, D. Walker, K. Hughes, and M. Weathers. "Preliminary Investigations of the Application of the Early Communication Indicator (ECI) for Infants and Toddlers." *Journal of Early Intervention,* vol. 28, no. 3, 2006, pp. 178–196.

Greenwood, C., S. Dunn,, S. Ward, and G. Luze. "The Early Communication Indicator (ECI) for Infants and Toddlers: What It Is, Where It's Been, and Where It Needs to Go." *The Behavior Analyst Today,* vol. 3, no. 4, 2003, pp. 383–388.

Greenwood, C., G. Luze,, G. Cline, S. Kuntz, and C. Leitschuh. "Developing a General Outcome Measure of Growth in Movement for Infants and Toddlers." *Topics in Early Childhood Special Education,* vol. 22, no. 3, 2002, pp. 143–157.

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.

Harjusola-Webb, Sanna M. "The Use of Naturalistic Communication Intervention with Young Children Who Have Developmental Disabilities." Ph.D. dissertation, University of Kansas, 2006.

Harms, Thelma, Richard Clifford, and Debby Cryer. *Early Childhood Environment Rating Scale—Revised Edition*. New York: Teachers College Press, 1998.

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

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: National Bureau of Economic Research, May 2013.

Heidemann, Sandra, Claire Chang, and Beth Menninga. "When Teachers Are Learning, Children Are Too: Teaching Teachers About Assessment." *Young Children,* 2005, pp. 86–92.

Hill, H. C., S. G. Schilling, and D. L. Ball. "Developing Measures of Teachers' Mathematics Knowledge for Teaching." *Elementary School Journal,* vol. 105, no. 1, 2004, pp. 11–30.

Hojnoski, R., B. Silberglitt, and R. Floyd. "Sensitivity to Growth Over Time of the Preschool Numeracy Indicators with a Sample of Preschoolers in Head Start." *School Psychology Review,* vol. 38, 2009, pp. 402–418.

Jenkins, J. R., J. J. Graff, and D. L. Miglioretti. "Estimating Reading Growth Using Intermittent CBM Progress Monitoring." *Exceptional Children,* vol. 75, 2009, pp. 151–163.

Kaminski, R. A., and R. H. Good. "Toward a Technology for Assessing Basic Early Literacy Skills." *School Psychology Review,* vol. 25, 1996, pp. 215–227.

La Paro, K. M., B. K. Hamre, and R. C. Pianta. *Classroom Assessment Scoring System (CLASS) Manual, Toddler*. Baltimore, MD: Paul H. Brookes Publishing Co., 2012.

Lambert, R., B. Rowland, H. Taylor, and C. Wheeler. *Resource Manual for Administrators and Principals Supervising and Evaluating Teachers of Young Children*. Technical report. Raleigh, NC: North Carolina Department of Public Instruction, 2010.

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

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.

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

Luze, Gayle, Deborah Linebarger, Charles Greenwood, Judith Carta, Dale Walker, Carol Leitschuh, and Jane Atwater. "Developing a General Outcome Measure of Growth in the Expressive Communication of Infants and Toddlers." *School Psychology Review,* vol. 30, no. 3, 2001, pp. 383–406.

Maheady, Larry, Michael Jabot, Janeil Rey, and Jean Michielli-Pendl. "An Early Field-Based Experience and Its Impact on Pre-Service Candidates' Teaching Practice and Their Pupils' Outcomes." *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children,* vol. 30, no. 1, winter 2007, pp. 24–33.

Malone, Lizabeth, Charlotte Cabili, Jamila Henderson, Andrea Mraz Esposito, Kathleen Coolahan, Juliette Henke, Subuhi Asheer, Meghan O'Toole, Sally Atkins-Burnett, and Kimberly Boller. "Compendium of Student, Teacher, and Classroom Measures Used in NCEE Evaluations of Educational Interventions. Volume II: Technical Details, Measure Profiles, and Glossary (Appendices A–G)." NCEE 2010–4013. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, April 2010.

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.

Maxwell, J. *Qualitative Research Design: An Interactive Approach*. London: Sage Publications. Thousand Oaks, 2011.

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

Meisels, S., and S. Atkins-Burnett. "Assessing Intellectual and Affective Development Before Age Three: A Perspective on Changing Practices." *The Food and Nutrition Bulletin,* special issue, 1999.

Meisels, S., S. Atkins-Burnett, Y. Xue, D. Bickel, S. Son, and J. Nicholson. "Creating a System of Accountability: The Impact of Instructional Assessment on Elementary Children's Achievement Test Scores." *Education Policy Analysis Archives,* vol. 11, 2003.

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.

Miles, M., and A. Huberman. *Qualitative Data Analysis: An Expanded Sourcebook*. Sage Publications, Inc., 1994.

Missal, K., J. Carta, S. McConnell, D. Walker, and C. Greenwood. "Using Individual Growth and Development Indicators to Measure Early Language and Literacy." *Infants and Young Children,* vol. 21, no. 3, 2008, pp. 241–253.

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 July 24, 2012.

National Center for Student Progress Monitoring. "Common Questions for Progress Monitoring." 2012. Available at http://studentprogress.org/progresmon.asp#2. Accessed March 15, 2013.

National Center on Quality Teaching and Learning. *Ongoing Child Assessment: Overview*. Available at http://eclkc.ohs.acf.hhs.gov/hslc/tta-system/teaching/docs/Ongoing-Assessment_TeacherTips.pdf. Accessed July 30, 2012.

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

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

Pianta, Robert C., Karen M. La Paro, and Bridget K. Hamre. "CLASS Classroom Assessment Scoring System Manual." Charlottesville, VA: Center for Advanced Study of Teaching and Learning, 2006.

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

Reschly, D., and J. Ysseldyke. "Paradigm Shift: The Past Is Not the Future." In *Best Practices in School Psychology IV*, vol. 1, pp. 3–20, edited by A. Thomas and J. Grimes. Bethesda, MD: National Association of School Psychologists, 2002.

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

Robson, C. *Real World Research: A Resource for Social Scientists and Practitioner-Researchers,* vol. 2. Oxford: Blackwell, 2002.

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

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

Safer, N., and S. Fleischman. "How Student Progress Monitoring Improves Instruction." *Educational Leadership,* vol. 62, 2005, pp. 81–83.

Shapiro, E. "From Research to Practice: Promoting Academic Competence Among At-Risk Students." *School Psychology Review,* vol. 37, 2008, pp. 46–51.

Shulman, L. "Those Who Understand: Knowledge Growth in Teaching." *Educational Researcher,* vol. 15, no. 2, 1986, pp. 4–14.

Smith, M., Joanne P. Brady, and Louisa Anastasopoulos. *User's Guide to the Early Language and Literacy Classroom Observation (ELLCO) Pre-K Tool.* Baltimore, MD: Brookes Publishing, 2008.

Speece, D., and L. Case. "Classification in Context: An Alternative Approach to Identifying Early Reading Disability." *Journal of Educational Psychology,* vol. 93, 2001, pp. 735–749.

Stuhlman, M. W., B. K. Hamre, J. T. Downer, and R. W. Pianta. "A Practitioner's Guide to Conducting Classroom Observations: What the Research Tells Us About Choosing and Using Observational Systems to Assess and Improve Teacher Effectiveness." Charlottesville, VA: The Center for Advanced Study of Teaching and Learning, University of Virginia, 2010.

Tilly, W. "Problem Solving Enterprise." In *Best Practices in School Psychology IV,* edited by A. Thomas and J. Grimes. Washington, DC: National Association of School Psychologists, 2002.

University of Oregon Center on Teaching and Learning. "UO DIBELS Data System." 2012. Available at http://dibels.uoregon.edu. Accessed March 15, 2013.

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. Office of Head Start, Administration for Children and Families, Department of Health and Human Services, January 2010. Available at http://eclkc.ohs.acf.hhs.gov/hslc/sr/quality/head_start_roadmap_to_excellence.pdf. Accessed October 31, 2013.

U.S. Department of Health and Human Services. "The Head Start Child Development and Early Learning Framework: Promoting Positive Outcomes in Early Childhood Programs Serving Children 3–5 Years Old." Washington, DC: U.S. Office of Head Start, Administration for Children and Families, Department of Health and Human Services, September 2011.

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.

Westen, D., and R. Rosenthal. "Quantifying Construct Validity: Two Simple Measures." *Journal of Personality and Social Psychology,* vol. 84, no. 3, 2003, pp. 608–618.

Whinnery, K., and P. Stecker. "Individual Progress Monitoring to Enhance Instructional Programs in Mathematics." *Preventing School Failure,* vol. 36, no. 2, 1992, pp. 26–29.

APPENDIX A

DEFINITIONS OF KEY DATA SOURCES AND METHODS

This page is left intentionally blank

DEFINITIONS OF KEY DATA SOURCES AND METHODS

Term	Definition
Data Sources	
Pedagogical content knowledge (PCK)	According to Shulman (1986), PCK includes knowledge of how to present content in a way that makes it comprehensible. PCK also includes knowledge of common student preconceptions or conceptions about content that help or hinder learning as well as knowledge about how to address any misconceptions. A measure of PCK should therefore capture the intersection of teachers' knowledge of child development, assessment, and instruction.
Scenario probe	In this context, a scenario probe involves short verbal or written vignettes with questions about decision-making/ processes designed to isolate what teachers know from the contextual circumstances that may affect what they do. Scenario probes may be embedded in a PCK measure.
Standardized pedagogical task (also called standard task)	In a standardized pedagogical task, teachers may be asked to review another teacher's documentation of a child, interpret the documentation, and use their interpretation to create an individualized plan for instruction. Teachers could also be asked to view a video of a classroom practice and record their responses, perhaps using an electronic tablet. Standard tasks may include a think-aloud protocol.
Think-aloud protocol	In a think-aloud protocol, teachers are asked to reflect on and discuss their use of ongoing assessment tools for instructional decision-making. These protocols draw on actual data from a teacher's classroom, often using a recent classroom observation or documentation as a basis for discussion. Teachers may also be asked to view a video of a classroom practice—either of themselves or another teacher—and discuss how assessment data were used to make decisions.
Methods	
Checklists	Evaluators use checklists to identify the presence or absence of behaviors, skills, or documents. Checklists should include clear definitions of each item being checked. For example, the types of ongoing assessment conducted—direct assessment, observation, and so on—could be items on the list. When accompanied by definitions, these items can be reliable and relatively objective measures; however, they are not appropriate for capturing gradations and qualitative content. Items on checklists can be accompanied by codes to help evaluators document evidence of certain content.
Ratings	Ratings are scales that take measurements along a continuum (for example, 1 to 5 or strongly disagree to strongly agree). 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 strongly do you agree/disagree with this statement: 'All of this teacher's documentation is objective'?"). Rating scales provide more detailed information than checklists and may be particularly useful for assessing quality and fidelity. However, the subjective nature of assigning ratings necessitates more training in order to obtain inter-rater reliability, compared with checklists.

Term	Definition
Rubrics	Rubrics are scoring guides that describe several levels of quality and multiple aspects of performance. They can be developed to document fine gradients of change to capture small but important differences across various aspects of quality. Rubrics incorporate ratings (for example, a scale of 1 to 5) but are more elaborate than ratings, often including multiple criteria that must be met before an item can earn points on the rating scale. *Analytic rubrics* are used to rate individual dimensions of quality (for example, to rate the flexible use of instructional strategies separately from reflection on the success of the strategies). *Holistic rubrics* are used to examine multiple dimensions or characteristics that co-occur (for example, to rate the flexible use of instructional strategies along with reflection on the success of the strategies and other aspects of individualizing instruction). Rubrics can be used for multidimensional concepts, such as examining how instruction is individualized to accommodate each child's strengths and weaknesses. 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 with 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 evaluator to complete than the other methods, which may be burdensome to teachers.

APPENDIX B

EXAMPLES FROM THE LITERATURE OF PRIOR USE OF DATA SOURCES

This page is left intentionally blank

EXAMPLES FROM THE LITERATURE OF PRIOR USE OF DATA SOURCES

Data Source[30]	Prior Use of Data Source
Document review and ratings	Fuchs et al. (1991b) used the Math-Modified Accuracy of Implementation Rating Scale–Revised at the elementary level to measure teachers' implementation of a curriculum-based measure that included technology-based support. The scale contains three subscales: (1) structure (estimating initial performance level, graphing scores, writing goals, drawing goal lines); (2) measurement (test administration and reliability of scoring); and (3) evaluation (describing instructional procedures and timing instructional adjustments). The study also used teachers' instructional plan sheets to examine their strategies for making instructional adjustments for each student. Teachers used these sheets to describe the details of each adjustment: (1) date, (2) instructional procedure, (3) arrangement, (4) time, (5) materials, and (6) motivational strategies. Goertz et al. (2009) collected a number of documents, including examples of teachers' classroom assessments and, when offered by the teacher, blinded samples of students' interim assessment work. A few teachers created their own templates for organizing data, and some shared copies with the researchers. The researchers used these documents to inform other data collection efforts, such as teacher interviews. Maheady et al. (2007) required teaching candidates to submit a written report about their experiences during an 8- to 10-week field placement in a preschool, primary school, or secondary school. The candidates were required to address five areas in their reports: (1) students and the educational context, (2) instructional goals and objectives, (3) assessment and instructional plans, (4) professional reflections and data analysis, and (5) intended professional responses.
Interview and observations	Goertz et al. (2009) used classroom observations and teacher interviews in the fall, winter, and spring to examine how elementary school teachers used interim and other formative assessments in mathematics to inform their instruction. Teacher interviews included semi-structured questions that provided context for the observed lessons and were designed to capture the ways teachers monitored students' mathematical understanding. During the winter interview, teachers were asked to "think aloud" about their classrooms' interim assessment results and to discuss patterns of class-wide performance and the math concepts with which their students struggled. To sort the interview data, the researchers used codes derived from a conceptual framework of teachers' use of assessment data to drive instruction. The data were sorted into descriptive categories that included three key themes: (1) the five primary domains that captured steps in the conceptual framework, (2) professional development and other available supports for teachers, and (3) the curriculum. Goertz and colleagues then used the coded interview data and classroom observation notes to create teacher profiles, a matrix that crossed steps in the instructional improvement cycle (data collection, interpretation, and action) with the type of formative assessment used.

[30] Data sources appear in the order they are introduced in the text.

Data Source[30]	Prior Use of Data Source
Teacher interview with think-aloud protocols	Using semistructured teacher interviews, Roehrig et al. (2008) asked elementary school teachers "to expand as much as possible on their experiences with assessment data so as to obtain a clear picture of their use of assessment data and how it influenced their instructional decision-making." The interviews had 28 open-ended questions, including inquiries about teachers' use of both assessment data and Florida's Progress Monitoring and Reporting Network (a delivery system for ongoing assessment data that pinpoints students who are not meeting expectations, identifies skills with which they are struggling, and offers recommendations for student grouping). Teachers were also asked about the barriers they encountered and the supports they received when using data to drive instruction. Using the grounded theory method, the researchers analyzed and grouped the interview data into open-coding categories and subcategories. They then identified emergent themes and developed a conceptual framework of teachers' use of assessment data to drive literacy instruction.
Pedagogical content knowledge with scenario probes	Goertz et al. (2009) used a survey to measure elementary school teachers' mathematical knowledge for teaching and to examine the relationship between their math knowledge and the ways they used information from interim math assessments. The survey had nine multiple-choice items on numbers and operations derived from the Content Knowledge for Teaching–Math instrument (Hill et al. 2004).
Standardized pedagogical task	In Goertz et al. (2009), the fall teacher interviews included a "data analysis scenario," during which researchers presented each teacher with a one-page printout of results from a hypothetical interim assessment. They asked each teacher: "to imagine that this was her class and to 'think aloud'…about what she saw in the results. After approximately five minutes, or after the teacher stopped talking, [researchers] continued with a series of six follow-up questions designed to call attention to patterns in the data ([for example], Are there any topics that this class, overall, appears to have difficulty with? How do you know?). In this way, [researchers] were able to capture . . . each teacher's initial, natural reaction to the assessment results as well as whether or not, with probing, she noticed particular strengths and weaknesses among her class" (p. 45). Teacher interviews in the fall and spring also included "misconception scenarios." In these scenarios, teachers received two prompts based on actual items from an interim assessment in which a fictional student's response showed an error or misconception. Teachers were asked (1) to offer a hypothesis identifying the student's misconception, (2) to imagine what question they would ask the student to verify that hypothesis, and (3) to describe how they would correct the student's misconception. The researchers used these data to construct a four-category typology of teachers' responses to student errors or misconceptions: (1) addressing student misconceptions, (2) engaging students in learning, (3) building on student ideas, and (4) promoting student thinking.

APPENDIX C

STEPS FOR PRE-TEST

This page is left intentionally blank

STEPS FOR PRE-TEST

Step 1: Recruitment and Consent

- Team recruits sites and collects class lists through a phone call with director and education coordinator.
  - Materials:
    - Contact information/lists
    - Advance letter
    - Call script
    - Classroom roster template
- Team ships consent packets to setting to distribute to two classrooms.
  - Materials:
    - Cover letters to setting
    - Classroom envelopes
      - Letter to teacher, two consent forms, envelope for lead teacher
      - Letter to family, two consent forms, envelope for each family
    - Shipping materials for the site to return the consent forms to Mathematica
- Team checks in on consent forms and schedules researchers' visit by phone.
    - In-person follow-up if needed
    - Materials:
      - Schedule
      - Extra consent packets
- Point person at the site ships consent forms to team.

Step 2: Preliminary Information

- Team ships OL&C to teacher.
  - Materials:
    - Cover letter for lead teacher
    - Packet of OL&C items for consented children in classroom (note: although it is not cost-effective for the pre-test, we could consider programming the OL&C into tablet for future data collection)
    - Express-mail return envelope for OL&C
- Teacher completes OL&C (approximately 15 minutes).
  - Returns via express mail
- Team ranks students and randomly selects one high performer, one student experiencing challenges, and alternates.
- Team ships to the teacher video equipment and a recording checklist explaining the desired recording sequence (method 1 or method 2).

- Materials:
  - Recording checklist
  - Tablet
  - Bendable mini-tripod
  - Video instructions

- Team communicates with teacher (by phone), identifying the two students selected and answering any questions (about 10 minutes).
- Teacher assembles documents for the two students in preparation for researcher visit (about one hour).

Step 3: Recordings with Tablet[31]

- Teacher makes plans to record target children.
- Teacher records a literacy or language assessment with at least one target child, using method 1 or 2.
- Teacher records small-group instructional activity related to literacy or language (involving two or more children) with at least one target child, using method 1 or 2.

Step 4: Researcher Visits

- Researcher visits classroom on scheduled day.
  - Materials:
    - Contact information and directions
    - Incentives and respondent payment receipts
    - Draft data collection manual (for example, item definitions and examples)
    - Document review instrument
    - Video observation instrument
    - Interview/think-aloud instrument
- Researcher reviews the documents and rates them.
- Researcher codes the video recorded by the teacher.
- Researcher prepares interview questions (influenced by gaps or questions from document review or video).
- Teacher participates in a one-hour interview with the researcher and completes a questionnaire.
- Researcher video- or audio-records interview with tablet for team review.
- Researcher provides the teacher with a gift card as a token of appreciation.
- Researcher finishes scoring all instruments.

[31] In the pre-test, we will ask teachers to describe the time required to do this.

Step 5: Debrief

- A team member calls a subset of the teachers to ask for feedback on the process, addressing items such as difficulty in assembling documents, ease of video recording, and length of each component (approximately 20 minutes).

This page is left intentionally blank

APPENDIX D

TEACHER QUESTIONNAIRE: ORAL LANGUAGE & COMPREHENSION

This page is left intentionally blank

TEACHER QUESTIONNAIRE: ORAL LANGUAGE & COMPREHENSION

DIRECTIONS: Consider the following statements as they relate to the child whose initials are listed at the top of the column.

Check the accompanying box in that child's column if you feel the statement is **TRUE** for that child.

ID:	**Child Initials**														
	___	___	___	___	___	___	___	___	___	___	___	___	___	___	___
1. This child often needs extra time in order to learn new things.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
2. This child often requires repeated instruction in order to learn a new skill.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
3. This child often needs to have activities simplified or modified in order to meaningfully participate.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
4. This child has trouble paying attention or staying engaged during large-group activities.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
5. This child has limited communication skills (e.g., limited use of nouns, verbs, adjectives, and adverbs when talking to you).	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
6. This child's behavior often interferes with his or her learning.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
7. This child is not able to participate independently at centers. He or she requires extensive supervision.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
8. This child does not know the names of common everyday objects, places, and things.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Source: Adapted from Bradfield, T.A., and S.R. McConnell. "Teacher Questionnaire of Language and Comprehension Skills Among Preschool Children." Minneapolis, MN: Center on Response to Intervention in Early Childhood, University of Minnesota, 2013. This is a research draft and may change. Prepared by Mathematica Policy Research.