



Early Childhood Teachers' Use of Ongoing Child Assessment to Individualize Instruction

OPRE Brief #2016-104

November 2016

A Brief Description Of The Examining Data Informing Teaching (EDIT) Measure

by Lauren Akers, Shannon Monahan, Sally Atkins-Burnett, Barbara Wasik, Judy Carta, and Felicia Hurwitz

This is one of a series of briefs about ongoing assessment for individualizing instruction.

This brief summarizes a measure of how teachers use ongoing assessment to individualize instruction.

An additional brief titled "Tailored Teaching: The Need for Stronger Evidence about Early Childhood Teachers' Use of Ongoing Assessment to Individualize Instruction" describes a review of the literature on ongoing assessment in early childhood settings for researchers and practitioners.

Finally, the brief "What Does it Mean to use Ongoing Assessment to Individualize Instruction in Early Childhood?" provides a conceptual framework that shows practitioners and researchers how practitioners can use ongoing assessment for individualization.

Using ongoing child assessment to individualize instruction is considered a best practice in early childhood educationⁱ and is a requirement in the Head Start Performance Standards (see Box 1).ⁱⁱ Teachers who use ongoing assessment to individualize instruction may deliver instruction that is more effective, reduce the school readiness gap for children at risk, and have students who achieve better outcomesⁱⁱⁱ Practitioners, researchers, and policymakers are paying increasing attention to how early childhood education teachers use ongoing child assessment to track children's progress and tailor instruction to each child's unique strengths, needs, and interests. In 2012, the Office of Planning, Research and Evaluation at the Administration for Children and Families (ACF) contracted Mathematica Policy Research and its partners to conduct a project titled "Assessing Early Childhood Teachers' Use of Child Progress Monitoring to Individualize Teaching Practices." The purpose of the project was twofold: (1) to develop a research-informed conceptual model for preschool teachers' use of ongoing assessment to individualize instruction, and (2) to create a measure, grounded in the quality constructs identified in the project's conceptual model, to examine this process (Box 2).

Box 1. Definitions of key terms

Ongoing child assessment: Repeated assessments and observations of a child's performance and progress over time. The most common ongoing assessment systems are closely tied to the curriculum and observations of children in learning activities.

Individualization: The process in which a teacher uses assessment or observation data to identify a child's skill level for a learning goal and tailor instruction for that child. The teacher uses data on an ongoing basis to see whether the child is progressing in response to the instructional changes and adjusts instruction as needed.

Box 2. Background information about the EDIT

To learn more about ongoing child assessment, the project's conceptual model, and the development of the EDIT, see ACF's website: <http://www.acf.hhs.gov/opre/research/project/early-childhood-teachers-use-of-progress-monitoring-to-individualize>.

This brief introduces the Examining Data Informing Teaching (EDIT) measure for researchers who seek to learn about preschool classroom teachers' use of ongoing assessment data to individualize instruction. In this brief, we discuss the EDIT's multi-method procedures, structure, and scoring; testing to date and future testing needs; a process for training EDIT raters; and potential uses of the measure.

OPRE

MATHEMATICA
Policy Research

KU

TEMPLE
UNIVERSITY
College of Education

Box 3. Curriculum-embedded approaches to ongoing child assessment

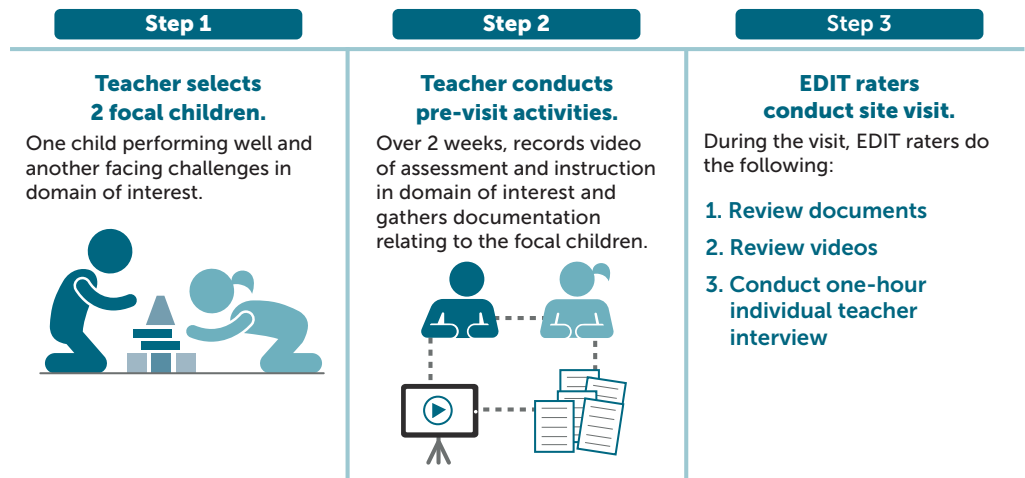
Curriculum-embedded approaches to ongoing child assessment draw on classroom activities as evidence of children’s performance and progress. Teachers use varied methods to collect data on children’s skills, knowledge, and behavior, including anecdotal notes, photographs, samples of work, videos, and completing ratings and checklists. The data are compared to behaviors described in rubrics that identify whether a child is below, at, or beyond expected level of performance for age. The EDIT focuses on curriculum-embedded approaches because they are (1) more common in early childhood settings than other ongoing assessment systems; (2) more demanding for a teacher to implement (that is, they require greater teacher skills and knowledge); and (3) more comprehensive, because they traditionally cover several domains of development.

The EDIT’s multi-method procedures, structure, and scoring

EDIT methods and procedures

The EDIT is a multi-method measure intended to assess preschool teachers’ use of curriculum-embedded approaches to ongoing child assessment and individualized instruction (Box 3). The EDIT consists of video-based observations, a document review, and a one-hour teacher interview (Figure 1).

Figure 1. EDIT procedures



- **Document review.** Teachers gather ongoing assessment data (such as a portfolio or anecdotal records) and lesson plans for two focal children, one performing well and the other facing challenges in a domain of interest (for example, language and literacy). At the site visit, EDIT raters examine the assessment data to see how teachers collect, organize, and use ongoing assessment data. Raters also review lesson plans and instructional materials for evidence of individualization.
- **Video-based observations.** Teachers video record a combination of assessments and instruction in the domain of interest with the two focal children, resulting in six observations collected across two weeks (Figure 2). During the site visit, EDIT raters watch the videos for evidence of how the teacher administers the assessments, and how the teacher modifies or individualizes instruction.

Figure 2. Focus and frequency of video-based observations

	Week 1		Week 2
Child Performing Well	Instruction in domain of interest (individual or small-group)	Assessment	Instruction in same domain (individual or small-group)
Child Facing Challenges	Instruction in domain of interest (individual or small-group)	Assessment	Instruction in same domain (individual or small-group)

- **One-hour semistructured individual teacher interview.** During site visit interviews, EDIT raters probe for additional explanations about the documents and video-based observations, as well as the teacher’s planning and implementation of adaptations, modifications, and individualized teaching strategies. Teachers describe how they use data to determine whether children are making adequate progress and any other ways that they use the data (for example, asking “How do you use the information you collect through observations or other ongoing assessments to inform what you do with children to help them learn?”).

EDIT structure

The EDIT measure consists of four holistic rubrics, three sets of analytic rubrics, five sets of ratings, and two checklists, for a total of 71 items (Table 1). The Appendix includes examples of different types of items. The EDIT captures the constructs in each stage of the process of using ongoing child assessment to individualize instruction; the development team selected the type of item (for example, checklist or rating) that best fit the construct within each stage, choosing the least burdensome type of item for raters when possible.

Table 1. Number and types of items, by process stage




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

Note: DR = document review; O = (video-based) observation; I = interview
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). *Ratings* are scales that take measurements along a continuum (for example, 1 to 4 or “not at all” to “always”). Raters use *checklists* to identify the presence or absence of behaviors, skills, or documents.

EDIT scoring

As depicted in Figure 3, each data collection method builds off the previous method for the raters to assign scores for the items.

Figure 3. Scoring the EDIT

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

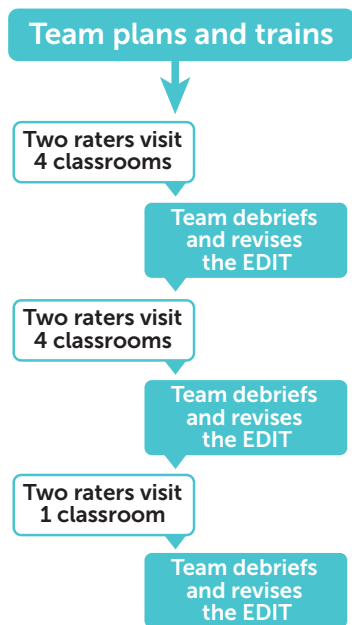
Testing to date

The development team pretested the EDIT to examine how the items performed and how our proposed procedures worked. We revised the items and procedures throughout the pretest.

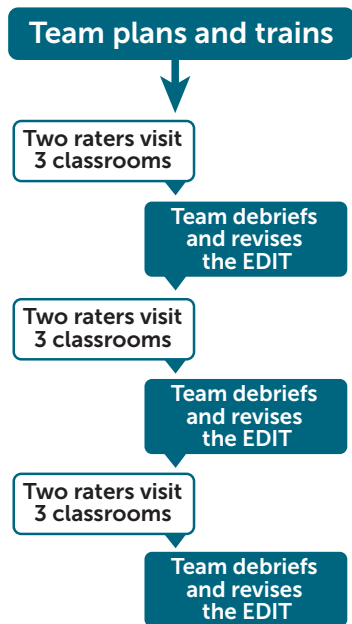
- Selection of centers and teacher
 - The iterative pretest consisted of two rounds of data collection (Figure 4).
 - The first round was conducted with nine English-speaking preschool teachers in six Head Start centers in three states.
 - The second round was conducted with nine English-speaking preschool teachers in seven centers (one of which was Head Start) in four states.
 - EDIT development team sought centers currently using a curriculum-embedded ongoing assessment system with moderate to high quality of implementation, drawing on recommendations from research or practitioner colleagues who had been in classrooms at the centers.
 - Supervisors at each center nominated teachers who collected and used ongoing assessment information regularly for the pretest. The teachers received a gift card as a token of appreciation for participation.
 - During the first round of visits, there was limited variation, with teachers and classrooms receiving high ratings on many of the initial EDIT items. After revising some EDIT items and visiting more classrooms, we found greater variability in the ratings during the second round of visits.
- Iterative revisions to the EDIT
 - After each set of visits, in consultation with ACF, the EDIT development team met and discussed challenges that arose and proposed changes to improve the procedures and EDIT items to obtain stronger measurement.

Figure 4. Iterative measure development

ROUND 1



ROUND 2



The development team then incorporated recommended changes and repeated the data collection and review process in the next set of classrooms as we iteratively refined the EDIT.

- Early during the first round of the pretest, the development team shared the EDIT with the project's expert panel, and incorporated members' feedback. We also shared findings with the expert panel between the first and second rounds and at the conclusion of the second round.
- Additional data collection during the pretest
 - At the end of the site visit, teachers completed a brief self-administered questionnaire (SAQ) to help the team learn more about teachers in the study. The SAQ included information on the teacher's demographics, early childhood background and experience, and the classroom.
 - After each round of data collection, the development team asked all teachers to participate in a 20-minute telephone call to debrief on the experience participating in the pretest, and most agreed to do so. The debrief discussions gave teachers a chance to comment on the process, including burden for collecting consents, gathering documents, and ease of video recording.
- Lessons learned from the pretest
 - Over the course of the pretest, the teacher's intentionality in data collection—and how this related to the teacher's use of ongoing assessment data to inform instruction—became a key focus of measurement. As such, the development team refined the criteria within rubrics and ratings, especially to capture measurement of teachers' intentional planning and evaluation of progress.
 - We changed some items to binary checklists (yes or no) or 4-point ratings rather than 7-point ratings in response to items with limited variation.
 - We added conventions to the instrument to support raters' completion of the instrument (for example adding "Not Applicable" as an option for some items that were situation specific).
 - We tested different ways of assessing constructs and selected the one that was most efficient and captured teacher practice.
 - During the debrief calls after the EDIT site visits, teachers reported that they would recommend participating in the study to colleagues. Some teachers reported that the project looked overwhelming when they first received our materials but then reported minimal burden (3 hours) after they got started. Teachers did not have difficulty operating the video equipment, and the resulting videos were of good audio and visual quality.

Next steps: Future testing needed

Only 18 classrooms were assessed with the EDIT during the development phase. More testing is needed with larger sample sizes to test the validity of the measure.

To further examine the psychometric properties of the EDIT, the development team recommends a larger pilot test with the following potential goals.

- Test the EDIT with a broader set of classrooms, looking for sensitivity to differences in practices and teacher backgrounds.
- Evaluate the reliability of the measure, including both internal consistency and inter-rater reliability.
- Evaluate different approaches to scoring the EDIT (for example, providing scores by dimensions and then averaging for a total score; weighting different items).
- Examine evidence of validity.
- Continue to expand operational definitions as needed to apply to a greater range in teacher practice.
- By conducting additional studies with much larger samples, raters could continue expanding training materials to address classrooms that use different assessment systems or teachers from differing backgrounds or who serve different populations.

Process for training EDIT raters

Rater training during the pretest

- Before our second round of visits, the development team trained two additional raters from outside the original EDIT development team. The training included item-by-item explanations (including more than 30 video and photograph examples gathered from previous visits), a description of the EDIT teacher interview, general site visit procedures, and a rating exercise. The two additional raters accompanied experienced EDIT development team members into the field, and came to consensus on ratings.
- For additional information on a recommended training plan, see Appendix E of the project's [final report](#). Appendix F of the final report provides training definitions and examples.

Future training recommendations

- For future uses of the measure, EDIT raters should receive training in how to review typical forms of documentation that teachers provide, how to review video-based observations of instruction and assessment, and how to conduct the EDIT teacher interview. Raters should also receive training in how to score EDIT items.
- Raters should receive training about the key features of the selected assessment systems and the terminology they may need to know to prompt teachers (during the interview) about their practices using a particular assessment. The rater needs to be familiar with the terms used in the

assessment to communicate more easily with the teacher about what the rater needs for review and tailor questions in the interview.

- The development team recommends sending two raters per classroom for a subset of classrooms for each assessment system, and examining inter-rater reliability based on independent ratings. After completing the visits, raters could discuss any discrepancies to determine whether they need additional clarifications to rate practices that a particular assessment system uses.
- The development team recommends conducting debriefing calls with raters weekly or biweekly, depending on the frequency of classrooms visits. These calls could help clarify unanticipated events that raters encountered in the field and provide additional information that will help refine wording and guidance in future training materials. For example, in order to ensure that raters understood items in a uniform way, based on these calls, we defined “measuring progress” as measurements at least three times within a reporting period. These calls could also ensure that raters are using the EDIT consistently and documenting the necessary information to justify their scores.

Potential uses after further research

After conducting additional research to more carefully examine the measure’s psychometric properties, the primary use of the EDIT would be to learn about what teachers are doing to collect and use assessment data to individualize instruction. The development team suggests the EDIT could be helpful in:

- Identifying beneficial practices for collecting and using ongoing assessment data to individualize instruction
- Informing professional development programs that support teachers in collecting and using ongoing assessment data to individualize instruction
- Examining and informing how programs support teachers in conducting and using assessment data

For more information on the EDIT measure, please refer to Monahan et al., 2016 on the Office of Planning, Research and Evaluation (OPRE) [website](#).

Appendix

This appendix provides sample items from the EDIT measure

Sample holistic rubric for “applying instructional decisions and individualizing”

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

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

Sample analytic rubric for “selecting and implementing the assessment method”

1	3	5	7	ENTER RATING
No evidence available OR The evidence collected is not aligned with assessment target(s).	The evidence collected is occasionally aligned with the assessment target(s). OR Evidence is very limited but is aligned.	The evidence collected is sometimes aligned with the assessment target(s).	Increases or varies The observed evidence collected is always aligned with the assessment target(s).	RATING: <hr/>

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

Sample analytic rubric for “selecting and implementing the assessment method”

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

	NOT AT ALL (ALMOST NEVER)	MINIMALLY CHARACTERISTIC (SOMETIMES EVIDENT)	STRONGLY CHARACTERISTIC (FREQUENTLY EVIDENT)	EXTREMELY CHARACTERISTIC (ALMOST ALWAYS EVIDENT)	CANNOT RATE
Teacher involves the family in interpreting and understanding the data.	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>	5 <input type="checkbox"/>

Sample checklist item for “organizing the information collected”

	YES	NO
The teacher views/organizes the data to compare a child’s performance to a developmental expectation or benchmark for growth.	1 <input type="checkbox"/>	2 <input type="checkbox"/>

Endnotes

- ⁱ National Association for the Education of Young Children (NAEYC). “NAEYC Early Childhood Program Standards.” Washington, DC: NAEYC, 2005. Available at <http://www.naeyc.org/files/naeyc/Position%20Statement%20EC%20Standards.pdf>. Accessed July 18, 2016.
- Sandall, S., M.E. McLean, and B.J. Smith. “DEC Recommended Practices in Early Intervention/Early Childhood Special Education.” Denver, CO: Division for Early Childhood of the Council for Exceptional Children, 2000.
- ⁱⁱ Administration for Children and Families. “Head Start Performance Standards.” 45 CFS Sec 1302.102.3(c)(2)(ii). *Federal Register*, September 1, 2016.
- ⁱⁱⁱ See, for example:
- Al Otaiba, Stephanie, Carol M. Connor, Jessica S. Folsom, Luana Greulich, Jane Meadows, and Zhi Li. “Assessment Data-Informed Guidance to Individualize Kindergarten Reading Instruction: Findings from a Cluster-Randomized Control Field Trial.” *Elementary School Journal*, vol. 111, no. 4, 2011, pp. 535–560.
- Buzhardt, Jay, Charles R. Greenwood, Dale Walker, Rawni Anderson, Waylon Howard, and Judith 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*, vol. 14, no. 3, 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.
- Fuchs, Lynn S., Douglas Fuchs, Carol L. Hamlett, and Pamela M. Stecker. “Effects of Curriculum-Based Measurement and Consultation on Teacher Planning and Student Achievement in Mathematics Operations.” *American Educational Research Journal*, vol. 28, no. 3, 1991, pp. 617–641.
- Landry, S., J. Anthony, P. Swank, and P. Monseque-Bailey. “Effectiveness of Comprehensive Professional Development for Teachers of At-Risk Preschoolers.” *Journal of Educational Psychology*, vol. 101, no. 2, 2009, pp. 448–465.
- Landry, Susan H., Paul R. Swank, Jason L. Anthony, and Michael A. Assel. “An Experimental Study Evaluating Professional Development Activities Within a State-Funded Pre-Kindergarten Program.” *Reading and Writing: An Interdisciplinary Journal*, vol. 24, no. 8, 2011, pp. 971–1010.
- ^{iv} The size of the subset will depend upon the size and expected variance in the sample and the availability of funding. At a minimum, inter-rater reliability should be conducted with each rater prior to the beginning of data collection and then at least once during data collection to check for rater drift. The lower the agreement between raters, the more paired observations are needed so that consensus scores can be used to improve the precision of the estimates. For more on considerations see Hallgren, Kevin A. “Computing inter-rater reliability for observational data: an overview and tutorial.” *Tutorials in quantitative methods for psychology* 8, no. 1 2012, p 23.