

EXECUTIVE SUMMARY

In 2012, the Office of Planning, Research, and Evaluation at the Administration for Children and Families (ACF) engaged Mathematica Policy Research and its partners to conduct a project titled “Assessing Early Childhood Teachers’ Use of Child Progress Monitoring to Individualize Teaching Practices.” The purpose of the project was twofold: (1) to develop a research-informed conceptual model for early childhood teachers’ use of ongoing assessment to individualize instruction, and (2) to create a measure to examine this process. Prior reports describe in detail the results of a literature review, conceptual framework, and measurement plan (Akers et al. 2014; Atkins-Burnett et al. 2014). This report describes the iterative development of the Examining Data Informing Teaching (EDIT) measure.¹ This report includes the results of a pretest study in 18 classrooms and a proposal for next steps for the EDIT.

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

The study involved six rounds of iterative pretesting for the EDIT measure. After each round, in consultation with ACF, the EDIT study team met and discussed challenges that arose and proposed changes to improve the procedures and EDIT items to obtain more precise measurement. We then incorporated recommended changes, and repeated the data collection and review process in the next set of classrooms as we iteratively refined the EDIT. Several times during the pretest, we shared the EDIT with the project’s expert panel, and incorporated their feedback.

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

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

Teachers in the pretest sample consistently received high scores on several EDIT items. We found evidence of some high quality general assessment practices that suggests it is relatively easy for teachers to score highly with certain types of items. For example, teachers regularly documented information about children objectively, and they collected the documentation during meaningful activities.² For some teachers, the electronic ongoing assessment systems (for example, TS GOLD) provided structured output that organized the data to make interpretation easier. Most teachers planned individualized lessons, using curricular materials aligned with the objectives that they assessed. During the first half of the pretest, these types of practices were indicative of the behavioral descriptions anchoring the high end of items and did not offer the opportunity to evaluate how well the rubrics worked for the middle of the scale to its lower end. We revised EDIT items for the second half of the pretest by adding measurement of intentionality and increased the criteria at the higher ends of the rating scales. These revised items resulted in greater variation in scores.

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

- **Intentionality.** Evidence of a child’s skill, knowledge, or behavior was limited in most areas to one or two observational records, and the method for collecting these records often was inefficient. The early versions of the EDIT did not capture the intentionality in data collection.
- **Instructional strategies.** Evidence of teacher awareness of how to identify evidence-based or professionally recommended instructional strategies was limited to the use of curricular materials. The strength of the link between child performance and the curricular activity was not reflected in the early versions of the EDIT.
- **Performance and progress.** When asked about using data for instructional decisions, teachers talked about performance rather than progress. They discussed where the children should be by the end of the year, but teachers during the first half of the pretest did not tell us how much progress children should be making in a given month or reporting period toward that end-of-year goal.
- **Planning and intentionality.** Evidence of planning and intentionality in collecting and using data to inform instruction and individualization was not adequately captured in the initial set of items.

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

- **Continued monitoring.** The initial set of items did not adequately capture continued monitoring of individualized goals or the review of the effectiveness of any instructional strategy or intervention. Only one teacher described using assessment data to determine whether or not an instructional approach was helping.

Throughout the pretest, we revised our data collection and administration procedures to ensure that we were able to capture the variation in teacher practice and validly represent how teachers use assessment. We changed the order of the questions and added prompts in the teacher interview to better obtain evidence of teachers' planning and evaluation of progress. We refined the wording on the rubrics to clarify and specify concepts, added indicators to better capture measurement of teachers' intentional planning and evaluation of progress, and added conventions to the instrument (such as "Not Applicable" rating options). We retained all three data sources (document review, video-recorded observations, and teacher interview). Each contributed some independent and some overlapping evidence about how teachers implement the process of using assessment to inform instruction and individualization. We expanded our measurement of planning, intentionality, progress, awareness of evidence based instructional strategies, and continued monitoring. With these modifications, we found greater variation in teacher's EDIT scores during the second half of the pretest (including evidence of intentionality, planning, and attention to progress within reporting periods).

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

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

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

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