Measuring Infant/Toddler Language Development: Lessons Learned About Assessment and Screening Tools

OPRE Brief

Yange Xue, Eileen Bandel, Cheri A. Vogel, and Kimberly Boller

Why Study Measures of Child Development?

Early Head Start programs are expected to assess child progress on an ongoing basis to ensure that children are developing the skills they need to be ready for school. Programs typically use two types of instruments to determine how children are doing: screeners and assessments. Screeners are typically used as a preliminary step to determine if developmental skills are progressing as expected or if skills are not progressing as expected and further evaluation is needed. Assessment tools are used on an ongoing basis to identify children's unique strengths and needs. Information from assessments can be used to determine strategies to support the development of the child within the context of the classroom. With a wide variety of assessment and screening tools available to select from, it can be daunting for Early Head Start programs to choose tools that will provide the information needed to understand children's progress and make decisions about instruction, further evaluation, and services. Early Head Start does not dictate which tools programs use; rather, programs select the assessments and screening tools they deem most appropriate for the population they serve. While this flexibility allows programs to tailor their practice, questions persist about how to select tools that are appropriate for the purpose and population at hand.

Crucial factors that programs must consider when selecting from among available tools include:

- qualifications of staff who administer the tool
- training requirements
- cost of administration
- ease of use
- purpose of the tool (screening vs. assessment)
- alignment with practice/curriculum
- appropriateness for the population
- reliability and validity of the tool

In many cases, these factors are at odds with each other, further complicating the selection process. For example, there are often trade-offs between a tool's reliability and validity and other considerations such as ease of use and administration cost.

MATHEMATICA
Policy Research

OPRE

Parent and staff reports are attractive choices because they are easy to use and are low in cost relative to standardized direct child assessments; however, compared to standardized assessments, little is known about their reliability and validity, particularly in a low-income Early Head Start population. Understanding the psychometric tradeoffs between tools can help program staff make the most informed decisions about further assessment and services.

This brief capitalizes on lessons learned in the Early Head Start Family and Child Experiences Survey (Baby FACES) to illustrate some of the considerations in selecting and administering screening and assessment tools in the context of Early Head Start. In particular, we focus on those tools that are used to examine the language skills of infants and toddlers and rely on parent and staff reports (rather than direct assessment).

While it is only one of five domains of child development (Box 1), we focus on language because it is foundational to children's learning across all five domains. Given its relative importance from birth to age three, this brief examines several parent- and staff-report tools that measure the language abilities of infants and toddlers. Although this brief is not meant to be a complete review of the wide array of language instruments that are available to screen and assess the infant and toddler population, we focus on several well-known parent- and staff-reported tools and observational measures used in Baby FACES. Specifically, we examine one developmental screener, the Ages & Stages Questionnaires Third Edition Communication scale (ASQ-3; Squires et al. 2009) and two language assessment tools, the MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al. 2000) and the Early Communication Indicator adapted for use in Baby FACES (ECI-Adapted; Carta et al. 2010; Luze et al. 2001). Drawing on information from Baby FACES, we ask and answer questions about how well these tools perform in a low-income Early Head Start population. The data from Baby FACES can provide information that might be useful to programs as they select and implement screenings and assessments in Early Head Start. Box 2 provides an overview of Early Head Start and the Baby FACES study.

Box 1. Domains of Child Development

The Head Start Child Development and Early Learning Framework defines five essential domains for infants and toddlers that align with and build from the school readiness domains identified by the National Education Goals Panel (Kagan et al. 1995). The domains include (1) language and literacy development, (2) cognition and general knowledge, (3) approaches toward learning, (4) physical well-being and motor development, and (5) social and emotional development.

What Are Reliability and Validity and Why Are They Important?

The reliability and validity of a tool are critical to determining its appropriateness. If a tool does not produce reliable or valid information, one cannot be confident that the resulting data reflect a child's true abilities. Box 3 describes the types of reliability and validity we examined using Baby FACES data.

If a tool is reliable, we can expect that the scores will be stable regardless of when the tool is administered, where it is administered, and who administers it. Reliability addresses the question: Is the tool producing consistent information across different circumstances? We evaluated internal consistency reliability by examining how closely responses to a tool's individual items are linked with each other. We assessed rater effects for staff-reported and observational tools by examining differences in scores across raters.

Box 2. About Early Head Start and Baby FACES

Early Head Start, created in 1995, is a two-generation program for low-income pregnant women and families with children up to age three that is administered by the Office of Head Start, Administration for Children and Families (ACF), U.S. Department of Health and Human Services. Early Head Start currently serves more than 100,000 children in about 1,000 programs, providing services primarily through child care centers, family child care homes, and home visits. Early Head Start programs provide comprehensive services including child development services, parenting education, case management, health care and referrals, and family support.

In 2007, the Office of Planning Research and Evaluation (OPRE) in ACF contracted with Mathematica Policy Research and its partners to conduct a descriptive study with a nationally representative sample of 89 Early Head Start programs (Vogel et al. 2011).[1] Baby FACES is designed to describe the national program and the experiences of two cohorts of children and families who were enrolled into the study in spring 2009: the Newborn Cohort consisted of pregnant women and newborns younger than 9 weeks old and the 1-year-old Cohort included children who were 10 to 15 months old at the time of the first data collection round. The study followed children, their parents, teachers and home visitors annually through their entire time in the program or until they turned 3.

[1] Does not include American Indian/Alaska Native programs Migrant/Seasonal Worker program

A valid tool measures what it is supposed to measure. A reliable tool produces the same results repeatedly. Both are needed for a tool to be useful.

If a tool is valid, we can expect that its scores accurately capture a child's progress on the developmental domain it is intended to measure. Validity addresses the question: Is the tool assessing what it is supposed to assess? To determine convergent and discriminant validity, we compared tools that have previously established reliability and validity information (the criterion tools listed in Box 4) to parent- and staff-reported tools and observational measures used in Baby FACES. Predictive validity was determined by linking parent- and staff-reported tools and observational measures to future performance on direct child assessments. Sensitivity and specificity of the tools were examined to determine the appropriateness of using them as a screening tool; we estimated the proportion of at-risk children (for sensitivity) and low-risk children (for specificity) defined by their scores on the criterion tools who were also correctly identified by the parent- or staff-reported tools or observational measures. We note here that infant and toddler development is inherently unpredictable and varying. During this early developmental period, skills do not develop in a linear fashion and young children's day-to-day responses to screening or assessment may vary greatly by their mood, time of day, and the circumstances surrounding the assessment. This poses challenges to monitoring infant and toddler development in a reliable and valid way.

Box 3. Types of Reliability and Validity Information Examined

- **Internal consistency reliability.** Internal consistency reliability demonstrates the extent to which a tool's individual items are related to one another. Cronbach's alpha is one of the most commonly used indicators of internal consistency reliability. The higher the value, the more related the items, and the higher the internal consistency reliability. Values of 0.70 or higher are usually considered to show acceptable internal consistency reliability.

- **Rater effects.** These are errors that occur when differences in scores on a tool are attributable to differences between the raters conducting the assessment rather than differences in children's skills. For example, some raters may tend to be more lax and some more stringent in their ratings, so that children's scores depend to some degree on who did the rating. Ideally, a child's score on a tool should truly reflect the child's performance. Rater effects are expressed by estimating the percentage of variability in scores that is attributable to the raters (that is, the intraclass correlation [ICC]). The higher the ICC, the less reliable the tool.

- **Concurrent convergent and discriminant validity.** Convergent and discriminant validity describe the degree to which a tool relates to other established tools in an expected way. With convergent validity, two tools measuring the same or similar things at the same time should be closely related to one another (such as two different measures of language development). Correlations of 0.50 or above on measures collected at the same point in time are considered to be strong evidence of convergent validity. With discriminant validity, two tools measuring different or dissimilar things at the same point in time should not be closely related (such as a measure of language and a measure of social-emotional development). Correlations below 0.30 on measures collected at the same point in time are considered good evidence of discriminant validity.

- **Predictive validity.** Predictive validity demonstrates the extent to which the tool of interest is associated with another tool administered at a later point in time. That is, predictive validity tells us about the extent to which scores collected at one point in time (say at age 1) are indicative of how the child will perform on another tool at age 2 or 3. A correlation between the two measures of about 0.30 or higher indicates good evidence of predictive validity.

- **Sensitivity and specificity.** Sensitivity and specificity are used to evaluate tools and refer to how well the tool correctly classifies and differentiates children who may have a developmental delay (sensitivity) and children who are developing normally (specificity). The higher the sensitivity and specificity of a tool, the greater its validity. Sensitivity and specificity are expressed as the proportion of children identified as at risk (or not) by both the tool of interest and by another tool measuring the same thing. Estimates of 90 percent or more are considered strong, 70 to 89 percent are moderate, and below 70 percent are low.

In the sections that follow, we describe each of the parent- and staff-reported tools and observational measures examined for evidence of reliability and validity: the ASQ-3 Communication scale, the CDI, and the ECI-Adapted. We then answer questions about how well these tools performed in a low-income Early Head Start population when compared to criterion tools that have previously established reliability and validity information (see Box 4). The findings are summarized in Table 1.

<div style="border: 1px solid blue;">

Box 4. Criterion Tools: Direct Assessments of Child Language Development and Assessor Ratings of Child Social-Emotional Development

- **Preschool Language Scale—Fourth Edition** (PLS-4; Zimmerman et al. 2002a, 2002b) is a direct child assessment used to evaluate receptive and expressive language skills, as well as understanding and use of grammatical rules for children from birth to 6 years of age. It is composed of two subscales: Auditory Comprehension (AC) and Expressive Communication (EC). We used the AC subscale at ages 2 and 3 in Baby FACES. The tasks in the AC subscale are designed to assess skills that are important for language development (such as following directions with cues and appropriate object play).

- **Peabody Picture Vocabulary Test, Fourth Edition** (PPVT-4; Dunn and Dunn 2007) is a norm-referenced standardized assessment designed as a measure of receptive vocabulary and verbal ability suitable for a wide range of ages, from 2½ through adulthood. In Baby FACES, we administered the PPVT-4 to all children at age 3 regardless of their primary language. Children are asked to say, or indicate by pointing, which of four pictures best shows the meaning of a word that is said aloud by the assessor.

- **Bayley Behavioral Rating Scale** (BRS; Bayley 1993) measures the child's behavior during child assessment. The BRS is one of the three component scales of the Bayley Scales of Infant Development—Second Edition (Bayley 1993). There are two subscales of the BRS used in Baby FACES:

  - **Orientation/Engagement** measures the child's cooperation with the assessor during the assessment, positive affect, and interest in the test materials.

  - **Emotional Regulation** measures the child's ability to change tasks and test materials, negative affect, and frustration with tasks during the assessment.

  The assessor rates the child's behavior by scoring items on a five-point scale, with 5 indicating more positive behavior (for example, more cooperation and less frustration). Scores are the total of the items in the subscale.

</div>

The Ages & Stages Questionnaires, Third Edition (ASQ-3)

The ASQ-3 (Squires et al. 2009) is a parent-report tool for screening infants and young children for developmental delays. Screening tools are designed to identify children who may be at risk for poor development and who may need special services and supports. Developer-derived cutoff points, which vary by age, indicate a need for further assessment.

The ASQ-3 includes 21 questionnaires that are appropriate for children aged 1 month through 5½ years. These surveys focus on the assessment of five key developmental areas: (1) Communication, (2) Gross Motor, (3) Fine Motor, (4) Personal-Social, and (5) Problem Solving. Parents reported on children's development at each wave of the Baby FACES data collection from age 1 to age 3. In this brief, we focus on the Communication score.

- **Does the ASQ-3 Communication scale measure children's communication skills consistently?** Yes. The Cronbach's alphas for ASQ-3 Communication scores were in the acceptable range, demonstrating adequate internal consistency reliability.

- **Does the ASQ-3 Communication scale measure what it is supposed to measure?** Yes, for the most part. ASQ-3 Communication scores were moderately correlated with two other language measures administered at the same age, the PLS-4 and PPVT-4, suggesting some evidence for

convergent validity. The ASQ-3 demonstrated some evidence of discriminant validity; correlations to the BRS Emotional Regulation and Orientation/Engagement scale scores were in the low range.

There is some evidence of predictive validity of the ASQ-3 Communication scores at age 1 for language outcomes at age 2 or age 3. The ASQ-3 Communication scores at age 1 significantly predicted PLS-4 scores at ages 2 and 3, but did not significantly predict PPVT-4 scores at age 3. By age 2, ASQ-3 Communication scores significantly predicted PLS-4 and PPVT-4 scores at age 3, suggesting stronger evidence of predictive validity than at age 1.

- **Does the ASQ-3 Communication scale identify children at-risk for possible language delays?** Yes and no—findings in Baby FACES were mixed. The sensitivity of the ASQ-3 as a screening tool for identifying children who may be at-risk for language delays was very low in the Baby FACES sample; this tool did not accurately identify children who were classified as at-risk based on their performance on the PLS-4 and PPVT-4. However, the specificity, a measure of correct exclusion from being identified as at-risk, was very high. Nearly all children who were identified as *not* at-risk for language delays (based on their PLS-4 and PPVT-4 performance) were correctly excluded from the at-risk category as defined by their ASQ-3 Communication scores. Taken together, the sensitivity and specificity findings suggest that the ASQ-3 may not be an appropriate tool for distinguishing Early Head Start children at-risk for language delays from those not at-risk.

MacArthur-Bates Communicative Development Inventories (CDI)

The CDI is designed to assess children's early communication skills. In Baby FACES, Early Head Start staff (teachers and home visitors) completed the English Infant Short Form for all children at age 1, the English Toddler Short Form for the 2-year-olds, and the CDI-III vocabulary checklist for the 3-year-olds. Staff reported whether the child understands (vocabulary comprehension) and says (vocabulary production) each of the words included in the checklist. In addition to staff reports, parents also reported on children's vocabulary production at ages 2 and 3.

- **Does the CDI measure children's language skills reliably and consistently?** Yes and no—findings in Baby FACES were mixed. Cronbach's alpha estimates for the parent- and staff-reported CDI were high, suggesting excellent internal consistency reliability. However, when looking a rater effects, the ICCs for the staff-reported CDI (with an exception for Vocabulary Production scores at age 1, where the ICC was lower) suggest that scores were dependent upon the staff members who made the ratings. This high rater bias in CDI scores decreases their reliability. The rater effects of the CDI varied, to some extent, by staff characteristics: educational level, whether they had a degree in early childhood education, and whether they were teachers or home visitors.

- **Does the CDI measure what it is supposed to measure?** Yes, except when looking at predictive validity of assessments at age 1. Both parent- and staff-reported CDI scores were correlated more strongly with direct language assessments than with the social-emotional comparison tools, suggesting some evidence of concurrent convergent and discriminant validity. Staff-reported CDI Vocabulary Production measured at age 1 significantly predicted PLS-4 scores at age 2, but did not significantly predict PLS-4 and PPVT-4 scores at age 3. Staff-reported CDI Vocabulary Comprehension at age 1 did not significantly predict PLS-4 and PPVT-4 scores at age 2 or age 3. These results suggest little evidence of predictive validity for the age 1 CDI. Parent- and staff-reported CDI scores at age 2 showed stronger evidence of validity, significantly predicting PLS-4 and PPVT-4 scores at age 3.

Table 1. Summary of Reliability and Validity Evidence for Parent/Staff Reported Tools and Observational Measures

| Tool | Assessment Mode | Reliability | | Validity | | |
		Internal Consistency Reliability (Cronbach's alpha)	Rater Effect	Concurrent Convergent and Discriminant Validity	Predictive Validity	Sensitivity and Specificity [a]
SCREENING TOOL						
ASQ-3 Communication	Parent report	**ACCEPTABLE** (0.65 to 0.82; most above 0.70)	n.a.	**MODERATE** Stronger correlations with PLS-4 and PPVT-4 scores (0.3 to 0.4) than with the BRS (0.2 to 0.3).	**AGE 1: WEAK** Explains 2 to 3 percent of the variance in PLS-4 scores at ages 2 and 3 **AGE 2: MODERATE** Explains 8 to 13 percent of variance in PLS-4 and PPVT-4 scores at age 3	**LOW SENSITIVITY** (< 10 percent) **HIGH SPECIFICITY** (> 96 percent)
ASSESSMENT TOOLS						
CDI	Parent report (only Vocabulary Production domain)	**EXCELLENT** (0.98)	n.a.	**MODERATE** Stronger correlations with PLS-4 and PPVT-4 scores (0.3 to 0.4) than with the BRS (0.2)	**AGE 2: MODERATE** Explains 10 to 14 percent of variance in PLS-4 and PPVT-4 scores at age 3	n.a.
CDI	Staff report	**EXCELLENT** (0.95 to 0.99)	**HIGH** ICCs: 49 to 63 percent. Lower (less biased) for staff with a BA degree or higher and those with an early childhood education degree; stronger rater effects for teachers (more biased) than for home visitors.	**MODERATE** Stronger correlations with PLS-4 and PPVT-4 scores (0.3 to 0.5) than with the BRS (0.2)	**AGE 1: WEAK** Explains less than 1 percent of variance in PLS-4 scores at age 2 **AGE 2: MODERATE** Explains 8 to 14 percent of variance in PLS-4 and PPVT-4 scores at age 3	n.a.
ECI-Adapted	Observation	n.a.	**LOW** ICCs less than 1 percent (except at age 2, where the ICC was 9 percent).	**WEAK** Weaker correlations with PLS-4 and PPVT-4 scores (< 0.2) than with the BRS (0.3).	**AGE 2: WEAK** Explains less than 2 percent of the variance in PLS-4 scores at age 3. No association to PPVT-4 scores at age 3.	**LOW SENSITIVITY** (< 50 percent) **LOW SPECIFICITY** (< 72 percent)

[a] Sensitivity refers to the proportion of children defined as at-risk for language delays (scoring one standard deviation or more below the mean) on the PLS-4 or PPVT-4 who were also identified as at-risk by the indicated tool (ASQ-3 Communication scale or ECI-Adapted). Specificity refers to the proportion of children identified as *not* at-risk (scoring higher than one standard deviation below the mean) on the PLS-4 or PPVT-4 who are correctly *excluded* from the at-risk category.

Early Communication Indicator-Adapted (ECI-Adapted)

The ECI-Adapted is a semi-structured, play-based assessment designed to measure the expressive communication of infants and toddlers between the ages of 6 and 36 months along four key skill elements: (1) gestures, (2) vocalizations, (3) single-word utterances, and (4) multiple-word utterances. Observations of these four elements are then combined to yield a total communication score. ECI was developed to be used to observe parents or a familiar caregiver playing with the child. For Baby FACES, we adapted this procedure and instead observed a play-based interaction between an assessor and the child.[1] Children scoring one standard deviation or more below the mean are at-risk for delays in expressive language (Greenwood et al. 2006, 2010). Though the ECI is an assessment tool, it is commonly used by practitioners to screen young children for possible language delays. Thus, in addition to examining the reliability and validity of the measure, we also look at sensitivity and specificity to understand whether it performs well as a screener.

Evidence of a tool's reliability and validity depends on the context in which the tool is used.

- **Does the ECI-Adapted measure children's communication skills reliably and consistently?** Yes. There was limited evidence of rater effects, suggesting that variation in children's scores is not attributable to differences in the performance of the raters.

- **Does the ECI-Adapted measure what it is supposed to measure?** No, there was little evidence for validity. Correlations between the ECI-Adapted scores and the PLS-4 and the PPVT-4 (administered at the same age) were in the low range. The ECI-Adapted scores correlated more strongly with the BRS Emotional Regulation and Orientation/Engagement scales than with the language measures, suggesting little evidence of convergent and discriminant validity. Evidence of predictive validity for the ECI-Adapted at age 2 is also limited; scores at age 2 significantly predicted age 3 PLS-4 scores but not PPVT-4 scores.

- **Does the ECI-Adapted identify children at-risk for possible language delays?** No. The sensitivity of the ECI-Adapted as a screening tool for identifying children with possible language delays was low in the Baby FACES sample; the tool did not accurately identify Early Head Start children who were classified as at-risk based on their PLS-4 and PPVT-4 scores. Specificity was also low, suggesting that the tool did not accurately exclude children who were *not* at-risk for language delays (based on their PLS-4 and PPVT-4 performance) from the at-risk category.

---

[1] We worked closely with the ECI developers to train and certify assessors and to standardize administration of the ECI as much as possible. Although parents would be familiar play partners to the child, we chose not to use them in this role because the task followed immediately after another parent-child interaction activity. Our concern was that parents would feel they were being corrected or admonished for their earlier interaction upon being instructed on how to "follow the child's lead" and limit the use of closed-ended questions during the play-based interaction (which is part of the ECI protocol).

What Should Programs Consider When Choosing an Assessment or Screening Tool?

The findings from analyses of the Baby FACES data suggest several implications for programs to consider when selecting measures of children's development.

Assessors need proper training and experience to administer measures validly.

Consider the purpose of the tool. Some of the tools used in Baby FACES did not perform well for their intended use. It is critical for programs not only to be mindful of the intended use of a tool, but also draw from available information (such as this brief) when selecting from among possible tools to use. For example, the Communication subscale of the ASQ-3 did not perform well as a developmental screener. When using this tool for this purpose and with a low-income population of infants and toddlers, the specified cutoffs may not be appropriate. For programs using screening tools with infants and toddlers, the weaknesses of the tool should be taken into consideration when making conclusions about further testing and service. Programs might also use results from more than one screener in conjunction to make decisions. However, programs may find that the ASQ-3 has other valuable uses; it may provide a vehicle for helping parents attend to developmental milestones and initiate conversations with program staff around these topics.

Use tools as they are designed to be used. Baby FACES invested in tools that were both new and promising and those that were well-established and commonly used by programs. Thus, data on the reliability and validity of Baby FACES measures can provide programs with information to use when selecting from among available tools. The experience with the ECI in particular showed the importance of adhering closely to the developers' intended administration. In our research-based administration of the ECI, we deviated from using a parent or other familiar play partner in favor of training and certifying assessors to act in this role. Children's relative unfamiliarity with the assessors may have led some children to communicate less than usual. It may also be that assessors (and those who later coded the activity) were less able to readily understand children's speech given their lack of familiarly with the children. Thus, the sample of child behavior and language captured by this tool may not have been representative of the child's typical communication, resulting in ECI scores that underrepresented children's true abilities.
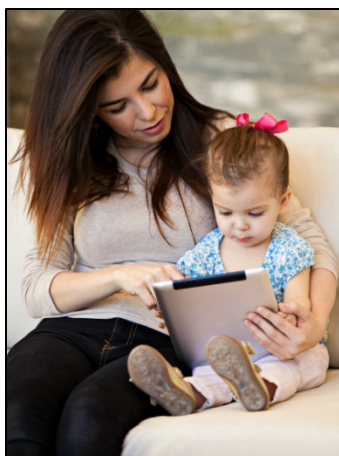
Not surprisingly, the ECI-Adapted showed limited validity in Baby FACES. These findings suggest that the ECI may not be well-suited for use in large-scale research studies (like Baby FACES) where such deviations in administration may be necessary. In addition, the tool showed little evidence that it can accurately screen children for possible language delays. However, programs may find this measure is well-suited for use as an ongoing progress-monitoring tool. While the ECI can provide a snapshot of children's communication proficiency at discrete periods during their first three years, the measure was designed to be administered on an ongoing basis to monitor the short-term growth and development of children's expressive communication.

Be aware that some tools require training to use properly. Another consideration when selecting the appropriate screener or assessment is knowing what training and experience is needed for the individual administering the tool. We found that a remarkable proportion of the variance in staff-reported tools of child language is due to rater effects: ratings of children's skills were

influenced by staff's training, education, and experience. Stronger rater effects were found for classroom teachers than for home visitors, suggesting that teachers may rate children's skills, knowledge, and behavior relative to other children in the classroom because, unlike home visitors, they regularly observe the children together with their peers. Rater effects were also found for staff with less than a bachelor's degree and staff without an early childhood education degree, suggesting that staff with more education and training in early childhood development are more likely to rate children relative to developmental norms. This raises concerns about the use of staff-reported tools for accountability purposes. It also suggests that to accurately measure children's skills, there is a need for specialized training of early care and education staff who perform child assessments.

Interpret results during the first year of life with caution. Across the board, we found that tools administered at age 1 were poor predictors of later development. This lack of predictive validity reflects a universal problem for assessments during infancy. During this early developmental period, skills are not sufficiently manifested to be measured with precision because development is not a smooth, linear process. By age 2, development begins to take a more linear form which is reflected in the current study's results. In contrast to findings for the tools at age 1, we found some evidence of concurrent and predictive validity for the parent- and staff-reported tools at ages 2 and 3. This suggests that these tools are adequate for measuring children's language skills during the toddler years.

The current work is not without its limitations. Although Baby FACES took great care to select child development measures that were appropriate for use with infants and toddlers, we operated within constraints. For example, we were able to include only a selection of the many tools available to measure infant/toddler language development and therefore, the findings reported here relate to a small subset of measures only in one developmental domain. Additionally, because Baby FACES is a study conducted with a low-income Early Head Start population, the findings may not be generalizable to other populations.

The findings from this study confirm that the reliability and validity of a tool needs to be considered in relation to the children being screened/assessed and the purpose of the screener/assessment. With an understanding of the tradeoffs between tools and the challenges to screening and assessing infants and toddlers, practitioners can make more informed decisions about further assessment and services. No single source of data will capture a full picture of young children's development. Parent- and staff-reported tools reflect children's skills and development in different contexts, which norm-referenced standardized tools otherwise cannot capture. Gathering information about children's development from multiple sources will help yield a more accurate picture of children's abilities and competencies.

References

Bayley, N. *Bayley Scales of Infant Development—Second Edition.* San Antonio, TX: The Psychological Corporation, 1993.

Carta, J.J., C.R. Greenwood, D. Walker, and J. Buzhardt. *Using IGDIs: Monitoring Progress and Improving Intervention for Infants and Young Children.* Baltimore, MD: Paul H. Brookes, 2010.

Dunn, L.M. and D.M. Dunn. *Peabody Picture Vocabulary Test—Fourth Edition.* Bloomington, MN: NCS Pearson, Inc., 2007.

Fenson, L., S. Pethick, C. Renda, J.L. Cox, P.S. Dale, and J.S. Reznick. "Short-Form Versions of the MacArthur Communicative Development Inventories." *Applied Psycholinguistics*, vol. 21, 2000, pp. 95–115.

Greenwood, C.R., D. Walker, and J. Buzhardt. "The Early Communication Indicator (ECI) for Infants and Toddlers: Growth Norms from Two States." *Journal of Early Intervention*, vol. 32, 2010, pp. 310–334.

Greenwood, C.R., J.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, 2006, pp. 178–196.

Kagan, S.L., E. Moore, and S. Bredekamp (Eds.). "Reconsidering Children's Early Development and Learning: Toward Shared Beliefs and Vocabulary." Washington, DC: National Education Goals Panel, 1995.

Luze, G.J., D.L. Linebarger, C.R. Greenwood, J.J. Carta, D. Walker, C. Leitschuh, and J.B. Atwater. "Developing a General Outcome Measure of Growth in the Expressive Communication of Infants and Toddlers." *School Psychology Review,* vol. 30, 2001, pp. 383–406.

Squires, J., E. Twombly, D. Bricker, and L. Potter. *ASQ-3 User's Guide.* Baltimore, MD: Paul H. Brookes, 2009.

Vogel, C.A., K. Boller, Y. Xue, R. Blair, N. Aikens, A. Burwick, Y. Shrago, B.L. Carlson, L. Kalb, L. Mendenko, J. Cannon, S. Harrington, and J. Stein. "Learning As We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children." OPRE Report 2011-7. Washington, DC: U.S. Department of Health and Human Services, Office of Planning, Research and Evaluation, Administration for Children and Families, 2011.

Zimmerman, I.L., V.G. Steiner, and R.E. Pond. *Preschool Language Scale, Fourth Edition (PLS-4).* San Antonio, TX: Pearson, 2002a.

Zimmerman, I.L., V.G. Steiner, and R.E. Pond. *Preschool Language Scale, Fourth Edition (PLS-4) Spanish Edition.* San Antonio, TX: Pearson, 2002b.

Suggested Citation

Xue,Y., E. Bandel, C.A. Vogel, and K. Boller (2015). Measuring Infant/Toddler Language Development: Lessons Learned about Assessment and Screening Tools. OPRE Brief 2015-52. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

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.