

**The Pew Home Visiting
Data for Performance
Initiative:
Phase II Final Report on
Parental Capacity and
Child Development
Indicators**

Deborah Daro, Sacha Klein,
and Tiffany Burkhardt

June 2017

ACKNOWLEDGEMENTS

This report builds on work initiated through the Pew Charitable Trusts' Home Visiting Campaign. The authors would like to thank the team at Pew who led this work—Karen Kavanaugh, the campaign director; Mimi Aledo-Sandoval, senior associate; and Danita Moses, administrative assistant—for their commitment to creating a common set of measures to assess home visiting performance. We also would like to thank Kay Johnson who contributed to the initial framework of the performance indicators and approach outlined in the initiative's Phase I report.

We are very grateful for the assistance and thoughtful comments of our technical advisory committee and researchers whom we interviewed as part of our review of the standardized measures and strategies used to assess parental capacity and child development. These individuals are listed in Appendices A and B. We are particularly appreciative of the input provided to the study by the national home visiting model representatives and state personnel administering funds from the Maternal, Infant and Early Childhood Home Visiting (MIECHV) initiative in their respective states. Improving the capacity of home visiting programs to capture change in parental capacity and child development in a more consistent manner is dependent upon the ongoing work of all national model and state leaders.

Finally, we want to acknowledge the continued financial support of our work by the Pew Charitable Trusts and the Trusts' continued commitment to improving the capacity of home visiting to support pregnant women, newborns and their families. The field has been well served by Pew's leadership in this area.

Deborah Daro, Chapin Hall at the University of Chicago
Sacha Klein, Michigan State University
Co-Project Directors

TABLE OF CONTENTS

Acknowledgements	1
Executive Summary	3
Overview.....	5
Context	10
Our Process.....	13
Parental Capacity Recommendations	16
Child Development Recommendations	21
Conclusion.....	33
Appendix A: List of Measures Examined	35
Appendix B: Key Informant Interviews.....	40
Appendix C: Advisory Board Meeting Attendees	41
Appendix D: Recommended Indicators	42

EXECUTIVE SUMMARY

The Pew Home Visiting Data for Performance Initiative (DPI) offers an important opportunity for states to galvanize around a small set of desired outcomes for the purposes of demonstrating the collective impact of investments in early home visiting. During Phase I, the initiative team, working in partnership with state leaders, national researchers, and several national home visiting models, identified a set of 11 indicators (nine which were defined at the end of Phase I and two additional indicators that were recommended for further study). Phase II of this work focused on pilot testing the feasibility of this proposed data collection approach and identifying ways to assess two additional outcome indicators of high interest to the early home visiting field: parental capacity and improvements in child development. This report summarizes the procedures and ultimate recommendations regarding outcome indicators in these two domains.

We began our work in both domains with a literature review of what measures are currently in use by home visiting programs and program evaluators as well as those measures cited by the research community as having strong utility in each domain. Over 30 child development measures and 35 parenting measures were examined and discussed with a group of external experts, state home visiting program directors, and representatives from several home visiting models, resulting in a narrowed list of 21 parenting measures and 12 child development measures which we explored further. Details for the narrowed list of measures, such as psychometric properties, sensitivity to change, cost, and length of administration, were researched and documented.

Key Findings

- Capturing change in caretakers' parental capacity and young children's development are not well served by administrative data sources and will require additional data collection by home visitors.
- In examining changes in parental capacity, it is more productive to measure parents' use of specific skills or parenting behaviors rather than shifts in parental attitudes.
- To the extent possible, multiple perspectives (i.e., parent self-report and home visitor observation) are needed to obtain an accurate assessment of how parents interact with their children and create safe and nurturing environments.
- Measuring changes in child development is complex, particularly for infants and toddlers younger than 18 months of age. We found no cost-effective, easy-to-administer direct measure that could reliably accomplish this task for this age group.
- Therefore, capturing the extent to which home visiting impacts development for children younger than 18 months old might best be achieved through an indirect measurement approach that documents the impacts of programs on parent sensitivity and responsiveness, a strong predictor of healthy child development.
- For programs that have the opportunity to observe children beyond 18 months of age, we do recommend direct measurement of child development, specifically in the areas of early language and social-emotional development.

- Regardless of the measures adopted by home visiting programs to monitor outcomes in the areas of parent capacity and child development, additional efforts are needed to improve staff training and the supervision of data collection to maximize data quality and reliability.
- Strategies to improve the quality and utility of assessing these domains, as with other outcome domains, is maximized when data are collected at the participant level, in its most basic, raw format, and at several points in time.

Specific Measurement Recommendations

Parental Capacity

We are recommending a suite of three brief instruments to serve as indicators for this domain.

Collectively, they capture dimensions of parental capacity from the perspective of program participants and home visitors, as well as a structured observation tool of parent-child interaction. These measures include:

- three of the nine subscales from the Healthy Families Parenting Inventory (HFI): participant self-assessment of her capacity to mobilize resources, parent-child behaviors, and quality of the home environment;
- the HOME-Short Form: completed by the home visitor to assess the caregiver’s parent capacity; and
- structured observation of parent-child interactions using either the PICCOLO or DANCE.

Child Development

Although promoting optimal child development is a critical component of home visiting, we are not recommending programs collect a direct measure of a child’s developmental progress for infants and toddlers under the age of 18 months. The variability and unevenness in how children develop in key areas related to later school achievement, such as language and socio-emotional development, during this period makes it difficult to draw reliable conclusions from repeated measures of these developmental outcomes. During this period, we instead recommend that programs carefully monitor the extent to which they are improving parent sensitivity and responsiveness through the repeated use of one of the Parental Capacity structured observation measures outlined above, as these are predictive of early childhood development.

For programs that have ongoing contact with children between the ages of 18 and 36 months, we do recommend programs collect direct measures of a child’s developmental progress, specifically the following assessments:

- MacArthur-Bates Communicative Development Inventories (CDI) Level II Short Form to assess progress in language development and
- The Brief Infant Toddler Social Emotional Assessment (BITSEA) to detect emerging social and emotional competence and problem behaviors.

Assessing change in children older than 36 months was not fully explored in our work in part because the majority of early home visiting programs focus their work on pregnant women and very young

children. As such, the measures we examined and the focus of our discussions centered on children 36 months or younger. Specific recommendations on appropriate tools for measuring developmental change for older children will require additional study.

In addition to these two measures, we recommend that home visiting programs enhance their use of the Ages and Stages Questionnaire (ASQ) and the ASQ-SE. While these are screening instruments and not typically used as change measures, preliminary research conducted by two members of the DPI-2 Advisory Board found evidence that programs may be able to adapt their use of these measures to assess the aggregate developmental performance of children participating in home visiting programs. By standardizing scores across multiple administrations of the tool, the DPI-2 research teams demonstrated that these tools may offer a promising path forward for monitoring program effects on the development of cohorts of young children served by early home visiting. Although ceiling effects do occur with the tool, in that a large proportion of children score at the highest levels of functioning and therefore lack room for measurable improvement, programs can address this issue by incorporating items in versions designed for older children. This approach would create expanded variance in the population-level results and improve the likelihood of identifying change over time at the program level. The use of this approach to assess change for individual children, however, was not examined and is not recommended at this time. Further testing on larger samples of home visiting participants, as well as data on a control or comparison group of children, would be required to determine if individual-level effects can be assessed using the ASQ-3 or ASQ-SE.

Enhancing Data Quality

Maximizing the utility of any of our recommended standardized measures across both domains will require programs to monitor the quality and consistency in how these tools are administered, analyzed, and used to inform ongoing service delivery. Initial training of staff and supervisors in the purpose of each tool and appropriate data collection methods is essential. Equally important, however, is offering ongoing support to home visitors in terms of supervising the data collection process, reviewing the results during supervisory sessions, and offering follow-up training to reinforce best practice. Program managers, working in partnership with staff, should routinely review trends emerging from these tools with respect to program outcomes and assess the implications of findings on the program's service delivery system.

OVERVIEW

The Pew Home Visiting Data for Performance Initiative (DPI) for Home Visiting offers an important opportunity for states to galvanize around a small set of desired outcomes for the purposes of demonstrating the collective impact of investments in early home visiting. During Phase I, the initiative team, working in partnership with state leaders, national researchers, and several national home visiting models, identified a set of 11 indicators (nine which were defined at the end of Phase I and two additional indicators that were recommended for further study) that states might adopt to determine whether their goals are being achieved across their full portfolio of home visiting programs, regardless of funding source. These indicators, summarized in Figure 1, focus on the more proximate and achievable outcomes likely to be realized through high-quality early childhood home visiting and build on data that states and programs are already collecting. To effectively monitor performance, they include a mixture of outcome measures and measures of processes used to reach outcomes. In addition, the recommended system includes 16 descriptive factors, such as several participant demographic characteristics, geographic location, and service delivery information, which states can use to more fully understand the differential impact of services on various subpopulations. The project also identified key strategies for enhancing the quality and usefulness of states' home visiting data—including collecting it in its most basic, raw format; at the participant level; and at multiple intervals—and comparing participant outcomes with those of similar families not receiving services. By increasing the rigor of their performance measurement and the utility of their data, states can strengthen the evidence base for home visiting, improve practice, and demonstrate its return on investment.¹

Figure 1.
Pew Home Visiting Data for Performance Initiative:
Phase I Recommended Indicators and Descriptive Factors

Maternal Health and Achievement Indicators

- Maternal Depression Screening and Referral
- Postpartum Health Care Visit
- Inter-birth Interval
- Maternal Educational Achievement

Child Health, Development, and Safety Indicators

- Child Development Screening and Referrals
- Child Development (further developed in Phase II)
- Well Child Visits
- Child Maltreatment
- Maternal Smoking and Tobacco Use

¹ The Pew Charitable Trusts (2015). *Using data to measure performance: A new framework for assessing the effectiveness of home visiting*. Philadelphia, PA: Author.

Figure 1, Continued

Parental Skills and Capacity Indicators

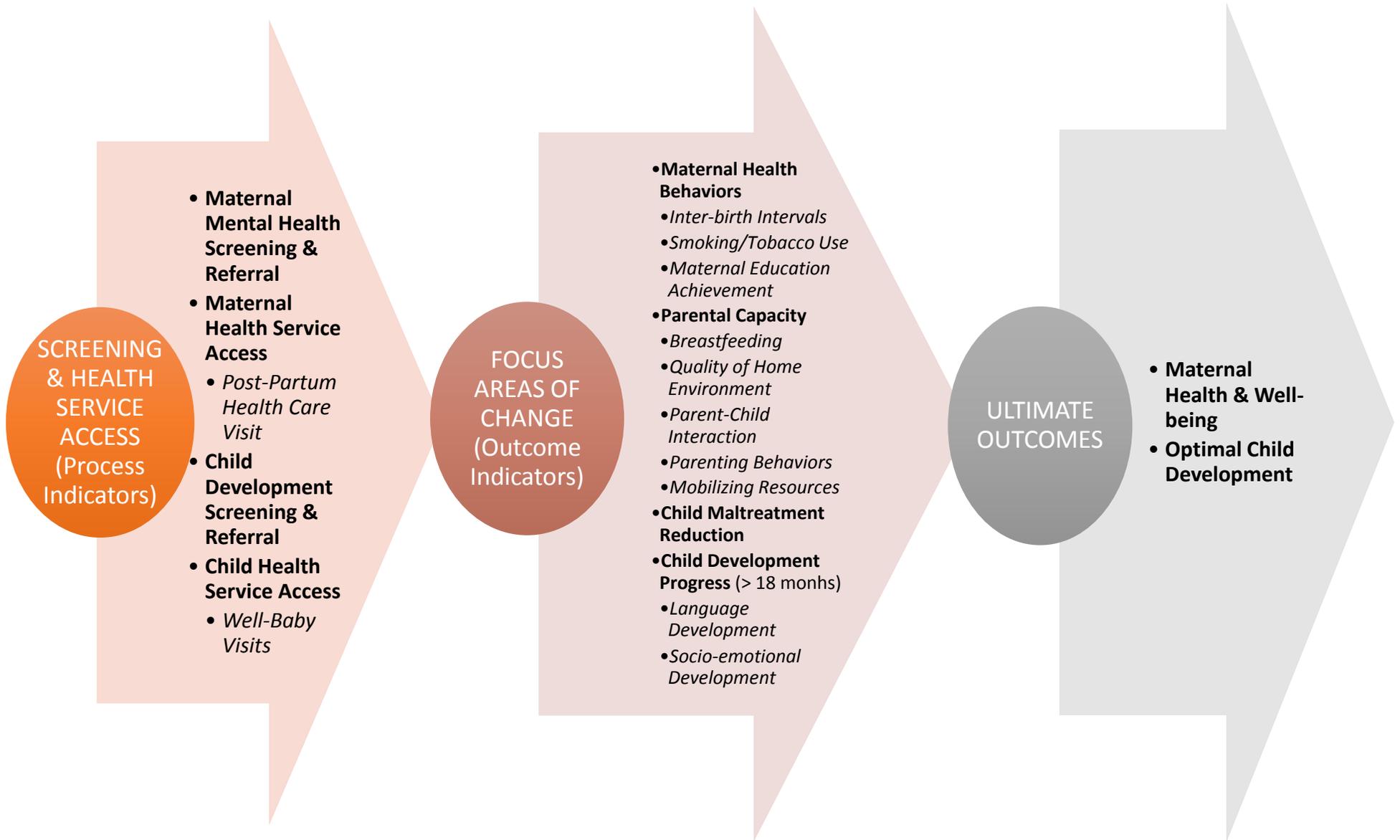
- Parental Capacity (further developed in Phase II)
- Breastfeeding

Descriptive Factors

- Child Characteristics
 - Date of birth
 - Full term versus preterm birth
- Maternal Characteristics
 - Date of birth
 - Number of prior births
 - Race/ethnicity
 - Native language
 - Whether the mom was in contact with the baby's father at time of enrollment
 - Residential zip code
- Participant Service Characteristics
 - Date of first home visit
 - Date of subsequent home visits
 - Date and reason for termination, including successful transitions and early terminations
- Program Data
 - Program model being implemented at site
 - Model expected number of home visits
 - Model supervisor-to-home visitor ratio
 - Average caseload per home visitor

In selecting our final list of indicators, the project was informed by a theory of change (see Figure 2) that placed primary emphasis on the role home visiting programs play in promoting maternal health and well-being and optimal child development. Although home visiting programs serve diverse populations and offer diverse content, early home visiting is distinct as an intervention strategy in its emphasis on strengthening parental capacity to address a child's immediate and ongoing developmental needs. The final list of indicators reflect our understanding on how home visitors utilize their relationship with their participants to achieve these early process and outcome indicators.

Figure 2. Pew Performance Indicators and Early Childhood Home Visiting Theory of Change



The Phase I efforts achieved significant progress in clarifying common outcomes and process indicators in multiple domains. However, fully operationalizing this system in any state requires additional information on how best to capture other important aspects of home visiting performance. It also requires identifying, specifically, how states will collect, manage, and use this information in real time. Phase II of the initiative, launched in February 2016, focuses on two specific activities:

1. Creating a shared understanding of how best to capture impacts in two additional performance areas—parental capacity and child development—central to the mission of nearly all early home visiting programs.
2. Pilot testing the proposed data collection system in a sample of states to determine the plan’s feasibility, identify key implementation challenges encountered by local home visiting programs or state agencies in coordinating their data collection efforts, and to test the extent to which administrative data matching techniques can be used to deepen the ability to track home visiting’s proximate and distal impacts while minimizing data collection burden for home visitors.

The purpose of this document is to report our recommendations related to the first of these two tasks. Specifically, the document outlines the process we followed in identifying measurement strategies in both of these high priority outcome domains—parental capacity and improvements in child development—and presents our recommended indicators and rationale. These additional measures, coupled with the work completed during Phase I, offers home visiting model developers, state administrators implementing these programs, and potential funders a solid system for measuring the collective impacts of early home visiting at the community, state, and, potentially, national level.

CONTEXT

State investments in home visiting are long standing. Over the past 40 years, a variety of funding streams have been used to support early home visiting programs. In some cases, states have drawn down Federal funding from maternal and child health care resources (such as Title X, Medicaid, and other health-related services), early education funding, and child abuse prevention dollars. In other cases, states developed their own early home visiting initiatives to support the replication of national home visiting models (such as the expansion of Healthy Families America in Florida, Indiana, and Illinois; the Nurse Family Partnership program in Louisiana and Pennsylvania; HIPPI in Arkansas; and Parents as Teachers in Missouri) or the development and dissemination of state-specific models such as the Nurturing Families Network in Connecticut and Follow Baby-Back Home in Arkansas. While these state efforts greatly expanded the availability of early home visiting in many communities between 1990 and 2010, the single biggest expansion of the program occurred as a result of the creation of the Maternal, Infant and Early Childhood Home Visiting (MIECHV) program, which was authorized as a part of the Affordable Care Act of 2010. This program allocated \$1.5 billion in formula and competitive grants to states to implement evidence-based home visiting programs during its initial five years, supporting local implementation sites in all 50 states and the US territories. Among the key features of the legislation is its requirement that states conduct comprehensive needs assessments to identify the communities most at-risk for poor maternal and child health outcomes; invest at least 75 percent of their resources in replicating specific evidence-based home visiting programs; and document the impact of these investments in altering performance in a set of benchmark indicators.

The initial Pew development work informed the U.S. Department of Health and Human Services Health Resources and Services Administration (HRSA's) work to revise the MIECHV benchmarks, which was finalized in March 2016. There is considerable overlap between the revised MIECHV benchmarks and Pew's home visiting indicators. However, the Pew effort focused on identifying the most concise list of indicator areas while the Federal system is more inclusive in identifying potential areas of program impact. In some domains of common interest there are substantial differences between how the revised MIECHV benchmarks and Pew measures are defined. For example, the one indicator area that Pew recommends states measure that is not included in any form in the revised MIECHV benchmarks is *inter-birth interval*. There are also five indicator areas included in the revised MIECHV benchmarks that are not included among Pew's list of home visiting indicators: preterm birth, safe sleep, child injury, intimate partner violence screening and referral, and continuity of health insurance coverage. The omission of these areas in the recommended Pew system does not indicate that these issues are not important outcome areas in which early home visiting may realize impacts. Rather, those working on the Pew initiative judged these and other candidate indicators as being less of a priority for including in Pew's robust but more succinct list of recommended indicators focusing on those outcomes currently targeted by the majority of home visiting programs.

The key differences between how the two systems recommend measuring the overlapping indicator areas are summarized below. A more detailed table comparing the revised MIECHV benchmarks and Pew's recommended home visiting indicators is available from the authors of this report upon request.

Primary Differences between Pew and MIECHV Indicators

Maternal Depression Screening and Referral

The measurement approaches are similar except the revised MIECHV benchmarks measure whether mothers with a positive depression screen receive at least one service contact, while Pew recommends measuring whether these mothers receive a service referral.

Postpartum Health Care Visit

The measurement approaches are almost identical except the revised MIECHV benchmarks measure whether mothers have a postpartum health care visit within 56 days of delivery and Pew recommends measuring whether mothers have a postpartum health care visit within 60 days of delivery.

Maternal Educational Achievement

The measurement approaches are almost identical except the revised MIECHV benchmarks measure this outcome for primary caregivers and Pew specifies measuring it for mothers.

Child Development Screening and Referral

The revised MIECHV benchmarks measure whether children with a positive screen for developmental delays receive services in a timely manner (which can include individualized developmental support from a home visitor, an evaluation by early intervention within 45 days of referral, or community services within 30 days), while Pew recommends measuring whether these children receive a service referral. Additionally, the revised MIECHV benchmarks measure receipt of developmental screens from any validated tool, while Pew specifies use of the Ages & Stages Questionnaire (ASQ).²

Child Maltreatment

The measurement approaches are very similar except the revised MIECHV benchmarks measure investigated child maltreatment reports whereas Pew recommends measuring all child maltreatment reports.

Well-child Visits

The measurement approaches are nearly identical. Both Pew and MIECHV recommend the AAP Bright Future schedule for well-child visits.

Maternal Smoking or Tobacco Use

The revised MIECHV benchmarks measure whether mothers who were smoking/using tobacco at enrollment are referred to smoking cessation services within 3 months (a process indicator), while Pew recommends measuring whether these mothers quit smoking (an outcome indicator).

² There has been some concern raised by states as to the recommended periodicity of this measure. The American Academy of Pediatrics (AAP) and the developers of the ASQ recommended slightly different time frames for administering the tool, which has resulted in variation in its use across home visiting models as well as across states. To date, the Pew initiative has not recommended a specific time frame.

Breastfeeding

The revised MIECHV benchmarks measure breastfeeding status when infants are 6 months, while Pew recommends measuring breastfeeding at 3 months.

Child Development

The revised MIECHV benchmarks measure suggests addressing this domain with two indicators: (1) the percent of home visits where primary caregivers were asked if they have any concerns regarding their child's development, behavior, or learning, and (2) the percent of children enrolled in home visiting with a family member who reported that during a typical week s/he read, told stories, and/or sang songs with their child daily. Both of these indicators are process measures that do not capture the extent to which the parental concerns or behaviors that are measured are indicative of children's actual developmental progress.

Parental Capacity

The revised MIECHV benchmarks address this domain by asking states to document the percentage of primary caregivers who receive an observation of caregiver-child interaction by the home visitor using a validated tool. Again, this is an indicator of program practice, but it does not capture the impact the practice might have on specific parent behaviors or attitudes.

Extending MIECHV Benchmarks in Important Domains

The revised MIECHV benchmarks reflect important progress in developing a more consistent and comprehensive picture of how early home visiting is achieving expectations, particularly in those domains of high priority to home visitors such as accessing maternal and infant health care, more timely screening and service referrals in the areas of child development and maternal depression, and altering maternal behaviors in areas such as breastfeeding and smoking. With respect to tracking the ability of early home visiting to influence child development and build parent capacity, the recommended MIECHV indicators underscore the difficulty in accurately capturing these two domains in ways that document meaningful change. Indeed, the DPI team spent considerable time during Phase I debating how best to measure both constructs but delayed recommending a specific approach pending further discussion.

Enhancing parental capacity and optimal child development are central to the mission of nearly all home visiting programs. As such, both are critical outcome domains and how programs measure them will influence how policymakers are likely to view the intervention's overall utility. Because of the importance of both of these areas, home visiting programs have developed multiple ways of defining and measuring these concepts, often in ways that are not fully complementary. No single measure reflects the breadth of either concept and simple, parsimonious measures are currently not available. Capturing progress in both domains will likely require home visitors to do additional data collection—no administrative data options exist with respect to reliable indicators of child development or parental capacity. All of these issues heightened the importance of our work in these domains and the challenges we anticipated facing as we moved forward.

OUR PROCESS

We began our work in both domains with a literature review of what measures are currently in use by home visiting programs and program evaluators as well as those measures cited by the research community as having strong utility in each domain. Details for each measure, such as psychometric properties, sensitivity to change, cost, and length of administration were researched and documented. A listing of the 30 child development measures and 35 parenting measures we examined is in Appendix A. After eliminating measures that would not be feasible to use in home visiting programs due to cost or complicated and lengthy administrative procedures, we conducted telephone interviews with researchers familiar with our candidate measures and with several instrument developers to explore the applicability and use of these measures within the context of early home visiting. A list of these key informants is provided in Appendix B. Specific areas we explored with these informants included each measure's scientific merits, ease of use, and relevance to program improvement. Drawing on the information provided by our key informants as well as an additional review of the literature recommended by those with whom we spoke, we created detailed tables outlining the relative advantages and limitations of each measures we considered.³ This process resulted in some measures being removed from the list and others being added.

After reviewing all of this input and conducting a final examination of the literature, we generated a final list of candidate measures: 12 child development measures and 21 parenting measures. Each of these measures was assessed in terms of ten criteria. The review criteria were: rigor (psychometrics and sensitivity to change), scope (comprehensiveness versus narrow aspect of the domain and applicability across early childhood, i.e., birth through five years), "fit" with home visiting culture (strength-based framework, available in multiple languages, capacity to improve program quality), and logistics (length, cost, ease of administration, and whether the measure was already widely used by home visiting programs). Summary descriptions of these criteria are in Figure 3.

To discuss this narrowed list of measures, we held a two-day review meeting with representatives of three key stakeholder groups—state home visiting leads, representatives from three of the largest national home visiting models (PAT, NFP, HFA), and researchers familiar with the measures and their application within home visiting programs. Attendees are listed in Appendix C.

While this meeting provided clarity on our approach to assessing parental capacity, appropriate measures with respect to the child development domain were more elusive. Specifically, we were unable to identify a standardized measure that would capture changes in a young child's development in the first year of life that was concise, reliable, and affordable. After extensive conversations with the Advisory Board, we elected to do a more in-depth review of a subset of standardized child development measures to determine if any of them or their key subscales were suitable for use with home visiting programs. Additional interviews were conducted with the developers and publishers of these measures.

³ These tables are available from the authors on request.

Using this additional information, we then convened the Advisory Board via conference call and formulated our final recommendations.

In the course of our discussions, the Advisory Board raised the possibility of exploring additional ways to use the current information a large proportion of home visiting programs are collecting through the repeated administration of the Ages and Stages Questionnaire (ASQ). This measure is commonly used as a screening tool designed to identify children at risk for developmental delays. It is not promoted as a measure of change. Two of the Advisory Board members with access to multiple ASQ assessments for a considerable number of children who had enrolled in home visiting in their respective states conducted secondary analysis of these data to determine the degree of change in these scores and in the proportion of children identified as being at risk for delays over time. In one of the studies, the researchers also had access to repeated ASQ scores for children not enrolled in home visiting. The results of these analyses were presented to the DPI study team in May and contributed to our final recommendations.

The balance of this report discusses the factors that contributed to our specific recommendations for parental capacity measures, approach to capturing changes in child development status, and strategies for insuring high-quality implementation.

Figure 3.
Summary Descriptions of Review Criteria

Factors	○	◐	●
Length	Long (30+ min)	Moderate (15-25 min)	Brief (< 15 min)
Cost	Expensive	Modest charge	Free
Sensitivity to Change Over Time	None/Minimal	Somewhat sensitive	Very sensitive
Program Value	Low value/interest for service providers	Moderate value/interest for service providers	High value/interest for service providers
Strengths-based Framework	Primarily captures changes in risk factors	Monitors a mix of risk and protective factors	Monitors primarily protective factors/strengths
Ease of Administration	Requires extensive training and monitoring to insure fidelity	Modest level of training and monitoring required	Minimal training/monitoring required
Psychometrics	Weak	Moderate	Strong
Comprehensiveness/Narrowness	Narrow	Moderate	Comprehensive
Already Widely Used by Programs	Rarely used	Used by two or more models	Used by multiple models
Availability in Multiple Languages	English only	English and Spanish	Multiple languages

PARENTAL CAPACITY RECOMMENDATIONS

Table 1 lists the 20 parental capacity measures we reviewed with our Advisory Board. In considering our recommendations in this domain, we paid particular attention to the importance of monitoring behaviors rather than attitudes. Many of the measurement and program experts we consulted with stressed the importance of tracking change in the frequency of specific behaviors as opposed to documenting attitudinal changes, noting that attending to behaviors or actions would yield a more robust and reliable estimate of program impacts. In addition, there was general consensus that it was important to obtain multiple perspectives on the nature and quality of the parent-child relationship. As such, we considered a range of strategies that would capture reliable provider feedback through direct observations of parent-child interactions as well as parent self-report on the frequency of certain behaviors in interactions with their child and their knowledge of child development. Finally, the feedback we received from both those we interviewed as well as the stakeholders attending our planning meeting emphasized the importance of capturing a variety of factors that impact a parent's ability to care for their child. A range of personal and contextual factors contribute to an individual's ability to meet the needs of their children. As such, it is important to monitor the potential program impacts from both the provider and participant perspective on such issues as the use of social supports and the quality of the home environment. In addition to obtaining the perspectives of home visitors and program participants through self-report measures and provider judgments, many of those with whom we spoke advocated for incorporating some form of structured observation of the parent-child interaction as a way of obtaining a more objective picture of parental capacity than can be achieved through provider assessment or participant self-report alone.

Based on this guidance, we are recommending a suite of indicators for home visiting programs to consider adopting to better monitor the ability of home visiting to improve or help parents build their capacity to meet their child's developmental and immediate needs. Collectively, these recommended measures capture a range of behaviors related to a parent's ability to manage their life and modify their behaviors in ways that protect their children, interact with their children in positive ways, and create an environment supportive of healthy child development. The advisory panel of experts that we convened strongly endorsed our recommending more than one measure to avoid constraining the science through overdependence on any single measure. The specific measures we are recommending include the following:

- **Healthy Families Parenting Inventory (HFPI):**

This tool would capture the parent's self-assessment of their parenting capacity. All parents participating in home visiting programs would be asked to complete three of the nine subscales from the HFPI (mobilizing resources, parent-child behavior, and quality of the home environment). Ideally, this self-assessment would be obtained at two points in time (intake and at a point close to program termination). All three scales have demonstrated a sensitivity to change over time with home visiting participant samples and the measure is free. Total time for parents to complete the three subscales: less than 10 minutes.

- **HOME-SF:**

This measure, a short form⁴ of the longer HOME Inventory, would capture the home visitor's assessment of a participating caregiver's parenting capacity. While the measure has been criticized by some as lacking sensitivity to cultural and SES/class differences and for some items being outdated, it has demonstrated sensitivity to change over time, including among low-income and culturally diverse populations. It is also widely used in the field and is familiar to many home visitors. Indeed, many of the experts we interviewed supported its use in terms of addressing a somewhat broader set of issues regarding the provider's assessment of the child's immediate environment. We contacted the instrument's developer (Robert Bradley) who is planning to revise the full HOME Inventory to address concerns regarding its unbiased application across diverse populations; however, the current version is still considered valid. The measure's overall familiarity among home visitors, its ease of administration, and planned improvements make it an attractive strategy of securing consistent provider assessment of this domain. Total time for home visitor to complete following a home visit: 5–10 minutes.

- **Observation of parent-child interaction using EITHER the PICCOLO or DANCE:**

In addition to recommending both a parent self-report measure and a provider assessment, we are also recommending that home visiting programs incorporate a parent-child observation measure. As discussed in the following section, carefully documenting how parents interact with their young children is an important vehicle for determining how early home visiting can improve parent sensitivity and responsiveness. While the HOME-SF has 8 observation items for ages 0 to 2 and 10 observation items for ages 3 to 5, the items measure whether providers observe any positive (or somewhat positive) parent-child interactions (e.g., mom kept child in view, mom did not spank the child, mom responded verbally to child, mom showed physical affection to child). In contrast, both the PICCOLO and DANCE assess the positivity, sensitivity, and responsiveness of the parent-child interaction in a more nuanced, comprehensive way. The nature of the items asked about and methods of scoring these more specific observation codes suggest a more accurate and consistent measure across individuals and a potentially higher sensitivity to change. More importantly, these two tools have been designed to inform practice in "real time." Although observational tools like these have historically not been adopted by most agencies because of cost and difficulty in administration, home visiting programs are increasingly embracing these strategies and are becoming more comfortable incorporating them into their standard practice precisely because of their instrumental value in improving practice.

We believe the potential strength of this strategy (as evidenced by the MIECHV recommendation that home visiting programs adopt them) provides confidence that such a tool

⁴ HOME-SF is available at the website for the National Longitudinal Survey of Youth (NLSY79):

<https://www.nlsinfo.org/content/cohorts/nlsy79-children/other-documentation/codebook-supplement/appendix-home-sf-scales/page/0/1/#AppendixA2A>

will, in the near term, be a common element in all home visiting programs. Indeed, both the PICCOLO and DANCE are currently being used in many states, suggesting that initial concerns as to the willingness or capacity of home visitors to integrate an observation tool into their practice may be less salient moving forward. The PICCOLO, limited to children ages 10 to 47 months, is widely used by several national home visiting models and several states are using this measure to meet MIECHV benchmark requirements. Lori Roggman, the developer of the PICCOLO, has conducted initial research applying the tool to infants 4 to 9 months. Researchers at Iowa State University also have collected observation data on infants enrolled in Iowa's MIECHV-funded home visiting programs suitable for additional validity testing. Initial indications from all these efforts are that the tool can be used reliably with infants.⁵ The DANCE is applicable for children from birth to 2 years old, and while it was originally designed for use only in NFP programs, it can now be used in other models. In addition, the developers of the DANCE are currently working on a distance learning approach for the DANCE designed to make the tool more accessible, feasible, and reduce costs for all home visiting programs.⁶

Detailed descriptions of the specific indicators captured in these strategies are outlined in Appendix D.

Advancing these Recommendations

Our general sense is that at least the first two of these recommendations could be adopted by state home visiting systems and programs rather quickly, at minimal cost, and with little disruption to the service delivery process. However, any widespread implementation of these recommendations requires attending to the following issues:

- Implementation of these recommendations would be substantially aided by support from the major home visiting models (such as NFP, HFA, and PAT). Toward this end, the national models, working collaboratively through the National Evidence-Based Home Visiting Alliance (the Alliance) should consider ways in which these measures might be adopted by all of their local affiliates. The eight national home visiting models currently collaborating in the Alliance (Child First, Family Connects, Healthy Families America, HIPPI, Nurse Family Partnership, Parent-Child Home Program, Parents as Teachers, and SafeCare) are among the most commonly implemented early home visiting programs across the country. They also reflect the diversity within the early home visiting field.
- To help ensure that home visiting systems are using culturally and socioeconomically competent measures, the Network of Infant/Toddler Researchers (NITR), a federally funded network of researchers focusing on early childhood, might offer one venue for partnering with Robert Bradley to develop and test an updated version of the HOME-SF that is more sensitive to culture and SES/class differences in home environments and parenting practices than the current version of this form. This update would also include modernizing some of the items, for

⁵ Personal communication with Lori Roggman, January 30, 2017.

⁶ Personal communication with Nancy Donelan-McCall, April 25, 2017.

instance, replacing the item asking about newspaper subscriptions with an item asking about using age-appropriate educational computer applications (apps).

- Using NITR and its members also may be helpful in exploring the feasibility of extending the use of the PICCOLO to a younger age group. Lori Roggman, the developer of the PICCOLO, is a member of NITR and one of several scholars participating in that group exploring the development of observational tools for infants. If Pew is interested in integrating an observational tool into standard practice within the context of existing home visiting programs, NITR may provide a productive platform for advancing this work.

Table 1. Parental Capacity Criteria Ratings

Note: Shaded measures were designed for use with a limited age range of children, while the unshaded measures are appropriate for families with children 0 to 5.

	Length	Cost	Sensitivity to Change Over Time	Program Value	Strengths-based Framework	Ease of Administration	Psychometrics	Comprehensiveness/Narrowness	Already Widely Used by Programs	Available in Multiple Languages
Assessment of Parenting Tool	●	●	--	●	◐	●	◐	◐	○	○
Dyadic Assessment of Naturalistic Caregiver-child Experiences (DANCE)	◐	○	●	●	◐	○	●	◐	○	○
Family Crisis Oriented Personal Evaluation Scale (F-COPES)	●	●	●	●	●	◐	●	◐	○	●
Family Map	○	◐	●	●	●	●	◐	●	●	◐
Family Assessment of Strengths Tool (FAST)	●	●	--	◐	●	●	◐	●	○	○
Healthy Families Parenting Inventory (HFPI)	◐	●	◐	●	●	●	●	●	●	◐
HOME-Short Form	●	◐	◐	●	○	●	◐	●	●	◐
Life Skills Progression	●	○	●	●	◐	●	●	●	●	◐
McMaster Family Assessment Device	●	●	●	●	◐	●	◐	●	○	◐
Parent Behavior Checklist	◐	●	●	◐	◐	●	●	◐	○	◐
Parent/Caregiver Involvement Scale (P/CIS)	◐	◐	--	◐	●	◐	◐	●	○	○
Parent Child Relationship Inventory	◐	○	◐	●	◐	●	◐	●	○	◐
Parent-Child Observation Guides (PCOG)	●	●	●	◐	●	○	◐	◐	○	◐
Parental Stress Scale	●	●	●	◐	◐	●	◐	◐	○	◐
Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO)	●	◐	●	●	●	○	●	◐	◐	◐
Parenting Scale	●	●	●	◐	◐	●	●	○	◐	◐
Parenting Sense of Competence Scale (PSOC)	●	●	◐	●	◐	●	●	◐	◐	◐
Parenting Stress Index – Short Form	●	○	●	◐	○	●	●	◐	◐	●
Psychological Empowerment Scale (PES)	●	●	●	◐	●	●	●	○	○	○
Support and Resiliency Inventory-Spouse	●	●	--	◐	●	●	◐	◐	○	○
3-Bag Task (Parent-Child Interaction Rating Scales)	◐	●	--	●	◐	○	●	◐	○	○

CHILD DEVELOPMENT RECOMMENDATIONS

Table 2 summarizes the 11 child development measures we discussed with our Advisory Board. In considering our recommendations in this domain, we paid particular attention to an instrument's ability to assess development for the full target age group (birth to 5 years old) served by early home visiting programs. Any assessment of child development, particularly early child development, is influenced by the very different capacities one might expect to see in children at different ages. As such, most of the measures we reviewed capture only a portion of the age range served by early home visiting programs. In certain cases, measures that are theoretically applicable to a broad age range use different questions or examples to assess children at different ages, complicating the ability to track developmental progress for individual children over time. At a minimum, it seems likely that separate tools may be needed for measuring gains in infant/toddler (birth to 2 years) and preschooler (3 to 5 years) development. In the course of our conversations, specific concern also was raised about the ability of any tool suitable for repeated administration by home visitors (as opposed to research staff) to accurately capture the developmental trajectory of a child under the age of one.

We also considered whether to capture child development broadly, across multiple domains (e.g., gross motor, fine motor, language, cognition, socio-emotional), or more narrowly. In contrast to our interest in capturing a number of different factors that shape parental capacity, a consensus developed from our interviews with key experts that child development measurement in home visiting would be best served, at this point in time, by focusing on measuring one or two specific aspects of early child development well as opposed to measuring more domains with less precision or at the risk of overburdening home visitors. The two areas that were repeatedly suggested by those we interviewed as offering the greatest promise for change as a result of early home visiting were child language development and socio-emotional development. While capturing both of these dimensions was considered an important objective for all home visiting programs, there was a slight preference, if a choice had to be made, toward assessing language development (both expressive and receptive language) over time. This was in part because of evidence that this developmental domain often predicts children's progress in other developmental domains, such as cognition⁷ and behavior.⁸

Instrument cost and training concerns are particularly daunting in this domain. The majority of psychometrically sound instruments used to assess a child's development are proprietary, very costly, time consuming to administer, and require substantial training and ongoing supervision to insure quality and reliability. Given these and other logistical constraints, the project team focused on identifying subscales from more comprehensive measures as the most viable option for capturing change in language development and socio-emotional development over time. In considering which measure and subscales to recommend, we paid particular attention to sensitivity to change (i.e., do these subscales reliably capture improvement or regression over time?); issues of cost (i.e., can costs be reduced if we

⁷ Marchman, V. A. & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, 11(3), F9–F16.

⁸ Petersen, I. T., Bates, J. E., & Staples, A. D. (2015). The role of language ability and self-regulation in the development of inattentive–hyperactive behavior problems. *Development and Psychopathology*, 27(1), 221–237.

use only certain subscales or might states be able to negotiate a reduced rate?); flexibility in administration (i.e., was a certain degree or professional experience required to administer the form in all cases?); and training and ongoing supervision issues (e.g., associated cost and time commitment).

Consideration also was given to the idea that the ASQ/ASQ-SE might be adapted for use as a change measure, which is an attractive idea from a feasibility standpoint because a large percentage of home visiting programs already use one or both of these instruments.⁹ It is not uncommon for program staff to use the ASQ as a way to document changes in a child's relative risk for developmental delays over time. However, these tools were designed for screening, not assessment, and are best used to identify children in need of further developmental assessment. Conventional wisdom suggests that using ASQ/ASQ-SE change scores as indicators of child developmental progress is problematic because of the high probability of ceiling effects and a lack of sensitivity to change over time, particularly for children who have developmental disabilities. Nevertheless, a consensus emerged among our Advisory Board members that the ASQ Communication Subscale and ASQ-SE may be useful indicators of developmental change for at least a subset of children receiving home visiting services, particularly if the focus was on measuring the percentage of children developing normally at program entry who are prevented from falling into the at-risk category.

After initially discussing available options with the Advisory Board, we were unable to identify a parsimonious solution that resolved all our concerns. As an alternative, we agreed to explore the utility of adopting specific subscales of more complex measures which focused on language development or social-emotional development to determine if this option made the administration of a measure of change in a child's development more feasible. The Advisory Board members recommended further conversations with the developers and publishers of five specific measures to assess the psychometric properties of administering individual subscales and the costs associated with this approach. These measures included the following.

- **Brief Infant Toddler Social Emotional Assessment (BITSEA)**

The Brief Infant Toddler Social Emotional Assessment (BITSEA)¹⁰ is the screener version of the longer ITSEA. Both are designed to detect emerging social and emotional competence as well as social-emotional and behavior problems and delays in the acquisition of competencies in children 12 to 36 months old. The BITSEA is a 42-item parent and caregiver report that takes approximately 7 to 10 minutes to complete, compared to 20 to 40 minutes required to complete the ITSEA. The BITSEA items were selected based on both clinical and empirical considerations, assessing externalizing behaviors (activity, aggression), internalizing behaviors (separation, depression), dysregulation (sleeping, eating), maladaptive habits, atypical, social relatedness, and competence (attention, compliance). The scale focuses on the development of competencies (e.g., hugs or feeds dolls or stuffed animals) as well as problem behaviors (e.g.,

⁹ Zaid, S. & Filene, J. (2014). *Design options for home visiting evaluation (DOHVE)- A DOHVE TA Resource Document*. OPRE Report #2014-56, Washington DC: Office of Planning, Research, and Evaluation, Administration of Children and Families, U.S. Department of Health and Human Services.

¹⁰ Carter, A.S., & Briggs-Gowan, M. (2005). *ITSEA BITSEA: The Infant-Toddler and Brief Infant Toddler Social Emotional Assessment*. San Antonio, TX: PsychCorp.

avoids physical contact). Respondents are asked to rate each item as “not true/rarely,” “somewhat true/sometimes,” or “very true/often.” The BITSEA was normed on a sample that was not nationally representative; the sample excluded children who, at birth, were expected to have severe developmental delays and excluded parents who could not speak English. However, reliability and validity are both in acceptable ranges, and the Advisory Board felt comfortable with the psychometric evidence for the BITSEA. Strengths of the BITSEA include that it is available in multiples languages, including English and Spanish, and can be administered by both parents and primary caregivers. This measure is widely used in early home visiting and other early intervention programs, and the BITSEA has preliminary support for sensitivity to change over time in Spanish.¹¹ In addition, the cost of the measure and the training requirements are both minimal.

- **MacArthur-Bates Communicative Development Inventories - Short Form**

The MacArthur-Bates Communicative Development Inventories (CDI) Short Form¹² was designed to measure language development in infants and toddlers. Three levels of the tool are available depending on the child’s age: 8 to 18 months for Infant form (Level I); 16 to 30 months for Toddler form (Level II); and 30 to 37 months for CDI-III (Level III). The Level I short form measures words and gestures and contains a vocabulary checklist of 89 words. Two versions of the Level II form both contain a productive vocabulary checklist of 100 words and a question about combining words. The Level III form contains the most items, comprised of 100 vocabulary words, 12 sentence pairs for assessing grammatical complexity, and 12 yes/no questions about semantics, pragmatics, and comprehension. All forms of the CDI are available in English and Spanish. Both receptive and expressive language scores are provided. Providers have found the CDI easy to administer, and parents generally find it useful. Reliability and validity have been found to be acceptable to high, but predictive validity is low at age 1 and moderate at age 2.¹³ The English versions of the CDIs were normed on approximately 1,800 children in three locations, and the Spanish versions were normed on more than 2,000 children. The long form has demonstrated sensitivity to change over time for children ages 16 to 30 months;¹⁴ however, more information is needed about the sensitivity to change over time for the short form. Discussions with the measure developers revealed concerns about using the CDI with infants. The developers would not recommend using the CDI as an outcome measure at 12 months and younger due to the floor effect with infants, and they generally advised against measuring individual change in language development with infants 12 months old and

¹¹ Hungerford, G. M., Garcia, D., & Bagner, D. M. (2015). Psychometric evaluation of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) in a predominately Hispanic, low-income sample. *Journal of Psychopathology and Behavioral Assessment*, 37(3), 493–503.

¹² Fenson, L. et al. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, 21(1), 95-116.

¹³ Vogel, C.A. et al. (2015). *Toddlers in Early Head Start: A portrait of 2-year-olds, their families, and the programs serving them*. OPRE Report #2015-10, Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

¹⁴ Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language & Cognitive Processes*, 12(5/6), 507–586.

younger.¹⁵ However, measuring language development over time with the CDI would be appropriate from 16 to 30 months of age with the Level II Toddler form and from 30 to 37 months of age with the Level III form.

- **Brigance Inventory of Early Development III**

This tool is a comprehensive assessment measuring six domains of child development containing both norm-referenced and criterion-referenced subscales. The Brigance is an appropriate tool for diagnostic assessment, progress monitoring, classroom assessment, and program evaluation. The third edition (IED-III, 2013) features expanded social-emotional and literacy domains. Reliability and validity are both adequate, with strong predictive validity for multiple child development outcomes. The IED-III was normed on a sample of over 2,000 children demographically representative of the US population, including English language learners. The tool is administered by direct assessment using the provided kit. Scores are recorded in the record book using the base and ceiling provided for each assessment to determine the first and last items to administer to each child. Two advantages of this tool are that (1) it includes subscales for communication and language as well as social-emotional development and (2) it is applicable for the full target age range (birth to 5 years). In fact, it can be used with children aged birth to 7 years, which would allow programs to use it as a follow-up measure should they choose to. The Language and Social-Emotional subscales were not designed to be administered independently, although others have done so in the past. Purchasing select subscales would not reduce the cost of the tool, but bulk discounts may be available depending on the volume ordered.

- **Vineland Adaptive Behavior Scales**

Useful for diagnosis, progress reporting, program and treatment planning, and research, the Vineland Adaptive Behavior Scales (VABS)¹⁶ are also applicable for the full target age range (birth to 5 years) of early childhood home visiting programs. Each item is rated on a 5-point scale, from Never (0) to Almost Always (4). Of the five domains included in this tool, we considered recommending the communication domain (receptive, expressive, and written subdomains). Online and paper administration and scoring are available. The Advisory Panel was aware of a study of children with autism that found the Vineland was sensitive to change over time, but more information about this consideration is needed. Domain-level versions of the forms are available in the third edition of this tool, providing domain scores and an overall adaptive behavior composite score. However, the measure developer confirmed that the domain-level version is only appropriate for children age 3 years and older because it was

¹⁵ Personal communication with Larry Fenson and Philip Dale, March 3-31, 2017.

¹⁶ Sparrow, S.S., Balla, D.A. & Cicchetti, D.V. (1984). *Vineland adaptive behavior scales*. Circle Pines, MN: American Guidance Service.

developed with the primary goal of conducting eligibility assessments, not as an outcome measure¹⁷.

- **Developmental Assessment of Young Children (DAYC-2)**

The DAYC was designed as an assessment to identify developmental delays, yet it can also be used to monitor child progress. The second edition (DAYC-2)¹⁸ was normed on a national sample of over 1,800 children representative of the US population. This measure includes both social-emotional and communication subscales, including subdomain scores for receptive and expressive language, for assessing children from birth through 5 years, 11 months. Each domain can be assessed independently by the examiner: standard scores, percentile ranks, and age equivalents are provided for each domain. Scoring is conducted by observing whether a child has mastered a skill or not. Scoring forms can be purchased separately for each domain, as well. This tool covers the full target age range, both developmental domains of interest, and is sensitive to change over time. However, the DAYC-2 uses pass/fail scoring where no credit is given for emerging skills, many materials are required that must be supplied by the examiner, and it is costly. Furthermore, the complex nature of calculating raw scores and converting scores to age equivalents requires considerable training.

The results of these discussions confirmed our initial observations on the difficulty of capturing change in a young child's language development and social-emotional development. In addition to technical challenges with these measures, such as the costs of purchasing the measures and insuring adequate staff training, the developers with whom we spoke raised theoretical concerns about trying to determine if changes observed in a young child's language development or social-emotional development before the age of 1 would be a valid and reliable indicator of program effects. Rather than assess changes in the child during this period, they recommended that programs focus on assessing parent sensitivity and responsiveness during infancy, an outcome many home visiting programs focus on and an outcome that is supported in the literature as a contextual condition for healthy child development.

An extensive body of literature supports the association between responsive parenting and adaptive social-emotional development in young children. Responsive parent-child interactions promote children's social, cognitive, and language development.¹⁹ Infants and children with responsive caregivers are less likely to exhibit behavioral and mental health problems and are more likely to be prosocial and succeed in school.²⁰ Maternal responsiveness has been linked with the development of emotion

¹⁷ Personal communication with Celine Saulnier, February 22, 2017.

¹⁸ Voress, J. K. & Maddox, T. (2013). *Developmental Assessment of Young Children, Second edition (DAYC-2)*. Austin, TX: PRO-ED.

¹⁹ See Table 1 in Roggman, L. A., Boyce, L. K., & Innocenti, M. S. (2008). *Developmental parenting: A guide for early childhood practitioners*. Baltimore, MD: Brookes Publishing.

²⁰ National Research Council, & Institute of Medicine (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. M. E. O'Connell, T. Boat, & K. E. Warner (Eds.). Washington, DC: National Academies Press.

regulation skills in early childhood²¹ and can even buffer the impact of genetic risk on emotion regulation.²² Synchronous, responsive interactions between the parent and child—sometimes referred to as “dyadic mutuality”—is key for the child’s emotional development and the parent-child relationship.²³ Dyadic mutuality in parent-child interactions is thought to create a sense of self-efficacy in the infant²⁴ and foster social,^{25,26} emotional,²⁷ and moral²⁸ development. Dyadic mutuality has been linked with fewer child behavior problems in older children²⁹ and effortful control in younger children,³⁰ which is related to self-regulated compliance.³¹ The importance of dyadic mutuality in the parent-infant interaction has been demonstrated in various populations, including high-risk, low-income families, as well as white, middle-class families.³² Findings from randomized controlled trials indicate that interventions that help parents increase the synchrony and responsiveness in their interactions with their children result in lower rates of disorganized attachment and more adaptive biobehavioral responses to stress.³³ Consistent responsive interactions with caregivers not only promote an emotional

²¹ Cole, P. M., Michel, M. K., & Teti, L. O. (1994). The development of emotion regulation and dysregulation: A clinical perspective. In N. A. Fox (Ed.), *The development of emotion regulation: Biological and behavioral considerations* (Vol. 59, pp. 73–100). Chicago, IL: Monographs of the Society for Research in Child Development.

²² Propper, C., Moore, G. A., Mills-Koonce, W. R., Halpern, C. T., Hill-Soderlund, A. L., Calkins, S. D., . . . Cox, M. (2008). Gene–environment contributions to the development of infant vagal reactivity: The interaction of dopamine and maternal sensitivity. *Child Development, 79*(5), 1377–1394.

²³ Stern, D. N. (1977). *The first relationship: Infant and mother*. Cambridge, MA: Harvard University Press.

²⁴ Brazelton, T. B., Koslowski, B., & Main, M. (1974). The origins of reciprocity: The early mother-infant interaction. In M. Lewis & L. A. Rosenblum (Eds.), *The effect of the infant on its caregiver*. (pp. 49–76). Oxford, England: Wiley-Interscience.

²⁵ Legerstee, M., & Varghese, J. (2001). The role of maternal affect mirroring on social expectancies in three-month-old infants. *Child Development, 72*(5), 1301.

²⁶ Tarabulsky, G. M., Tessier, R., & Kappas, A. (1996). Contingency detection and the contingent organization of behavior in interactions: Implications for socioemotional development in infancy. *Psychological Bulletin, 120*, 25–41.

²⁷ Kochanska, G., & Coy, K. C. (2002). Child emotionality and maternal responsiveness as predictors of reunion behaviors in the strange situation: Links mediated and unmediated by separation distress. *Child Development, 73*(1), 228.

²⁸ Kochanska, G. (2002). Mutually responsive orientation between mothers and their young children: A context for the early development of conscience. *Current Directions in Psychological Science, 11*(6), 191–195.

²⁹ Deater-Deckard, K., & Petrill, S. A. (2004). Parent-child dyadic mutuality and child behavior problems: An investigation of gene-environment processes. *Journal of Child Psychology and Psychiatry, 45*(6), 1171–1179.

³⁰ Kim, S., & Kochanska, G. (2012). Child temperament moderates effects of parent-child mutuality on self-regulation: A relationship-based path for emotionally negative infants. *Child Development, 83*(4), 1275–1289.

³¹ Kopp, C. B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology, 18*(2), 199–214.

³² Raver, C. C. (2004). Placing emotional self-regulation in sociocultural and socioeconomic contexts. *Child Development, 75*(2), 346–353.

³³ e.g., Bernard K., Dozier M., Bick J., & Gordon M. K. (2015). Intervening to enhance cortisol regulation among children at risk for neglect: Results of a randomized clinical trial. *Development and Psychopathology, 27*, 829–841.

connection between the caregiver and child, but also stimulates the child's cognitive, social, emotional, and physical growth^{34,35}.

The "give and take" of contingent responsiveness is particularly important for language development.³⁶ Parent-child engagement through responsive and positive play promotes children's language development.^{37,38,39} A meta-analysis of 18 studies of parent-implemented language interventions indicated that parents receiving the interventions were more responsive and used more language models, both of which promote child language development.⁴⁰ Children with responsive caregivers in early childhood are also more likely to succeed in school.⁴¹ Children's social-emotional competence and language development are facilitated by responsive parent-child interactions that include reciprocity and engagement through play and conversation.

Given the difficulty in reliably capturing developmental changes in children under one, we are not recommending home visiting programs implement new measures to directly assess child development during infancy. The recommendation is not a reflection of the value we place on this domain. Rather, it is a reflection of the current state of measurement development and the best time point for capturing these changes. The majority of early home visiting programs initially engage families when a woman is pregnant or when a child is born. As noted above, research on the early development of a child's social-emotional well-being consistently reports the importance of parental sensitivity and responsiveness in fostering a context for healthy development. As such, we believe that carefully documenting change in

³⁴ Institute of Medicine and National Research Council; Committee on the Science of Children Birth to Age 8: Deepening and Broadening the Foundation for Success; Board on Children, Youth, and Families. (2015). *Transforming the workforce for children birth through age 8: A unifying foundation*. L. Allen & B. B. Kelly (Eds.). Washington, DC: The National Academies Press.

³⁵ National Research Council and Institute of Medicine; Committee on Integrating the Science of Early Childhood Development; Board on Children, Youth, and Families; Commission on Behavioral and Social Sciences and Education. (2000). *From neurons to neighborhoods: The science of early childhood development*. J. P. Shonkoff & D. A. Phillips (Eds.). Washington, DC: National Academy Press.

³⁶ Gadsden, V. L., Ford, M., & Breiner, H. (Eds.). (2016). *Parenting matters: Supporting parents of children ages 0–8*. Washington, DC: National Academies Press.

³⁷ Leffel, K. & Suskind, D. (2013). Parent-directed approaches to enrich the early language environments of children living in poverty. *Seminars in Speech and Language, 34*(4), 267–277.

³⁸ Sheffield Morris, A., Robinson, L. R., Hays-Grudo, J., Claussen, A. H., Hartwig, S. A., & Treat, A. E. (2017). Targeting parenting in early childhood: A public health approach to improve outcomes for children living in poverty. *Child Development, 88*(2), 388–397.

³⁹ Weisberg, D. S., Zosh, J. M., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Talking it up: Play, language development, and the role of adult support. *American Journal of Play, 6*, 39–54.

⁴⁰ Roberts, M. Y. & Kaiser, A. P. (2011). The effectiveness of parent-implemented language interventions: A meta-analysis. *American Journal of Speech-Language Pathology, 20*, 180–199.

⁴¹ National Research Council, & Institute of Medicine (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. M. E. O'Connell, T. Boat, & K. E. Warner (Eds.). Washington, DC: National Academies Press.

parental practices and the quality of the parent-child relationship during a child's first 18 months of life offers the best early indication of the contribution home visiting makes to healthy child development.

However, for programs that enroll children later or retain children in services beyond infancy, assessing change in development would be advisable. For home visiting programs that have the opportunity to observe children beyond 18 months of age, we recommend that direct measures of child development be adopted. Because of the importance of early language development and social-emotional development for later developmental outcomes and educational performance, we would recommend measuring language and social-emotional development in children ages 18 to 36 months. Specifically, we are recommending the MacArthur-Bates CDI (Level II or Level III short form, depending on the child's age) and BITSEA. The child development measure recommendation would apply to programs that serve families with children in this age range, regardless of the age of child at enrollment. For these child development measures, the preferred schedule would be to administer the CDI and BITSEA at 18 months, 24 months, 30 months, and 36 months of age, if the child is enrolled in the program during this time.

Assessing change in children older than 36 months was not fully explored in our work in part because the majority of early home visiting programs focus their work on pregnant women and very young children. As such, the measures we examined and the focus of our discussions centered on children 36 months or younger. Specific recommendations on appropriate tools for measuring developmental change for older children will require additional study. A detailed description of suggested measures for children ages 18 to 36 months is included in Appendix D.

Examining the ASQ as a Measure of Change

While our initial review of the ASQ-3 and ASQ-SE suggested that these screening tools were not appropriate to use as developmental change measures, two members of the Advisory Board (David Bard at the University of Oklahoma Health Science Center⁴² and Craig LeCroy at Arizona State University⁴³) conducted exploratory research on the potential of the ASQ-3 to be used in this way, particularly for those children who are developing normally when they enter the program. The almost universal use of the ASQ across all home visiting models and its similarity in structure to the ASQ-SE made this possibility particularly attractive. If these measures could be adapted to serve as an indicator of change, even at the aggregate level, the majority of home visiting programs would be in alignment in assessing some common domains related to optimal child development within adopting a new measurement tool.

The ASQ-3 is comprised of five main subscales evaluating development in Communication, Gross Motor, Fine Motor, Problem Solving, and Personal-Social Abilities. Twenty different versions of the tool have been developed, each of which target children at a specific point between the ages of 2 months and 60 months. Each version has the same subscales and scoring range (0 to 60). Although items do appear

⁴² Bard, D. & Hunter, M. (2017). *Squeezing developmental change out of the ASQ3 scores: A report on child development outcomes for the Home Visiting for Performance initiative*. Available from Chapin Hall on request.

⁴³ LeCroy & Milligan Associates, Inc. (2017). *Options for measuring changes over time on the ASQ-3*. Available from Chapin Hall on request.

across more than one version of the tool, the content of each version varies to reflect the different skills and competencies children would be expected to demonstrate as they age.

The two DPI research teams conducted secondary data analyses on multiple ASQ3s obtained from children enrolled in early home visiting programs in their respective states and, in the case of the Oklahoma study, a mostly unserved community sample (84% self-reported no current or past use of home visiting). The Arizona sample used ASQ-3 scores from 215 children assessed at 6, 9, and 12 months of age who were participating in Health Families America home visiting programs. The Oklahoma sample included data from 2010 to the present on 4,509 children for whom one or more ASQ was completed between the ages of 2 and 60 months as part of participation in one of four home visiting programs (Nurse Family Partnership, Healthy Families America, Parents as Teachers, and SafeCare). An additional sample of 1,488 children with ASQ data who had participated in a community research program also was included in the Oklahoma sample. Item-level information was only available from the community sample, so the Oklahoma psychometric investigation focused on this largely unserved group of children ranging in age from 2 to 60 months. Comparisons of subscale scores were also conducted, examining patterns separately for each home visiting sample and for the community sample.

The primary research question posed by both studies was: Can the ASQ-3 as implemented by home visiting programs be used to measure change in child development over time? Strategies employed by one or both teams included examining: (a) trends using existing scoring methods and categories (typical, questionable, or delayed); (b) trends using existing raw scores; (c) trends using standardized scores to facilitate comparison across different versions of the ASQ; (d) trends examining only the common questions that appear on two or more of the ASQ versions; (e) the correlations observed between general summative questions (e.g., does anything about your baby worry you?) and the total ASQ score; and (f) the statistical fit and theoretical plausibility of a common factor (for each subscale) explaining variation across items of all ASQ versions. All of the analyses conducted by both teams focused on the performance of the tool as a measure of change on aggregate samples (and subsamples) as opposed to determining the tool's ability to detect developmental change over time for an individual child (i.e., neither team attempted to develop new normative scoring metrics of developmental change at the individual child level).

Both teams found the ASQ useful as an evaluation tool for tracking developmental change over time at the program level. By rescaling the scores across all versions, the research teams demonstrated that the subscales of the individual versions of the tool can be reliably linked to provide a picture of how cohorts of children moving through a home visiting program develop over time. While the individual teams used different statistical methods in building their arguments, both demonstrated that the concept of using the ASQ as a change measure was feasible and appropriate. More detailed descriptions of each team's analytic methods can be found in their respective reports, available from Chapin Hall or the authors directly upon request.

These studies also highlighted the need for additional research to more fully understand how home visiting programs might interpret any changes observed over time in the aggregate performance of the children they serve. While program managers can observe the magnitude of change in their participant

population over time, additional research, involving a comparison or control group, is needed to determine if a given level of change is meaningful. Also, additional research is needed to more fully understand how much time is needed between assessment periods to obtain a robust indicator of change. Test-retest estimates were excellent with children after 12 months. For younger children, it is unclear if lower test-retest reliability estimates reflect difficulties in capturing certain constructs, such as communication for very young children, or if changes in the first 12 months are simply less predictable (abilities change rapidly and cannot be forecasted well when solely relying on past assessments), supporting our initial concerns with home visiting programs attempting to directly measure (or at least interpret) developmental changes during this period.

Since the vast majority of children in both samples are not identified with delays, and many children achieve perfect scores during the assessments, the ASQ, as currently administered, suffers from a ceiling effect in that children with the most positive scores have very little room to demonstrate improvement when completing subsequent versions. The research teams suggested that home visiting programs consider adding items from the adjacent versions of the ASQ to their earlier assessments (e.g., for the 3-month version of the ASQ, include items from the 6-month and 12-month version) to increase variability in responses and limit the number of times a parent responds “yes” to every question. Home visitors can then use these items with parents to highlight developmental changes they might be expected to observe over the next few months and provide guidance on how to nurture these skills with their children. Similar guidance also could be given to parents on how to appropriately address a specific delay their child may demonstrate. Independent of the ASQ’s measurement capabilities, the ASQ captures many of the common developmental milestones included in more complex developmental assessments. It offers home visitors a useful tool in communicating these milestones to parents and encouraging parents to be more intentional in how they promote their child’s development. Its repeated use also may sensitize parents to early indicators of potential problems. Indeed, one of the findings noted by the Arizona team was the high correlation between a parent’s response to a general question about her child’s development (e.g., does anything about your baby worry you?) and actual ASQ scores. This finding supports the decision to include, as part of the MIECHV benchmarks, a standard that encourages home visitors to explicitly ask caretakers if they have any developmental concerns regarding their child.

Advancing these Recommendations

We are not recommending home visiting programs adopt direct measures of child development if the majority of children they serve exit the program before 18 months of age. Both our initial review of various standardized measures and our assessment of the ASQ-3 suggests early home visiting programs should place highest priority on implementing a consistent and reliable strategy for conducting multiple observations of parent-child interactions and documenting progress in improving parent sensitivity and responsiveness during a child’s first year. The parent observation tools recommended for the parent capacity domain can be used to capture these concepts. For home visiting programs that enroll children beyond 18 months of age, we are recommending the repeated use of CDI and BITSEA for children between 18 and 36 months.

To advance a program's ability to directly measure changes in child development, we believe home visiting programs would be well served by conducting more rigorous assessments of their ASQ-3 and, potentially, ASQ-SE scores, paying particular attention to aggregate changes observed in the children they serve over time. The types of rescaling methods employed by the DPI advisory board members offer programs an additional tool for exploring the direct impacts home visiting programs might have on early child development. Consistent use of the tool may enhance a parent's understanding of upcoming developmental milestones and strengthen their capacity to more fully support their child's optimal development.

As with our recommendations in the parenting domain, widespread adoption of these strategies will require the involvement and support of all key stakeholders in the home visiting field, including model developers, state home visiting lead agencies, Federal MIECHV administrators, and members of the research community. We are encouraged by the expanded use of parent observation tools (particularly the PICCOLO) among MIECHV-funded home visiting programs, which suggests the adaptation of such measures will become more widespread.

Table 2. Child Development Criteria Ratings

Note: Shaded measures were designed for use with a limited age range of children, while the unshaded measures are appropriate for families with children 0 to 5.

	Length	Cost	Sensitivity to Change Over Time	Program Value	Strengths-based Framework	Ease of Administration	Psychometrics	Comprehensiveness/Narrowness	Already Widely Used by Programs	Availability in Multiple Languages
Ages and Stages Questionnaire (ASQ - 3)	●	●	--	◐	◐	●	◐	●	●	●
Behavior Assessment System for Children (BASC-3)	◐	○	●	◐	◐	◐	◐	●	○	◐
Behavior Problems Index	●	●	●	◐	○	●	◐	○	○	◐
Brigance Inventory of Early Development III	○	○	●	◐	●	○	●	●	○	◐
Developmental Assessment of Young Children (DAYC-2), 2nd edition	○	○	●	●	●	○	◐	●	○	○
Developmental Observation Checklist System (DOCS)	○	○	--	◐	●	◐	●	●	○	○
Early Learning Accomplishment Profile (E-LAP)	○	○	--	●	●	○	●	●	○	◐
Infant Toddler Social Emotional Assessment (ITSEA/BITSEA)	○/●	○/●	●	●	◐	●	●	◐	◐	●
MacArthur-Bates Communicative Development Inventories- Short Form	●	◐	●	●	●	●	●	○	●	◐
Parents' Evaluation of Developmental Status (PEDS)	●	◐	●	◐	○	●	◐	◐	○	◐
Social Skills Improvement System (formerly Social Skills Rating System)- Social Skills Scale	◐	○	●	●	●	●	●	○	○	◐
Vineland Adaptive Behavior Scales; Social-Emotional Early Childhood Scales	○/●	○	--	●	◐	◐	●	●/○	○	◐

CONCLUSION

The Pew Home Visiting Data for Performance Initiative is designed to offer a starting point for building a consistent and well-researched approach to evaluate early home visiting as a strategy, irrespective of model or funding stream. We are recommending an observational parenting measure—either the PICCOLO or DANCE—and the select HFPI subscales and HOME-SF to measure parenting capacity. To measure child development, we are recommending the CDI and BITSEA only for children 18 to 36 months of age. The child development measure recommendation would apply to programs that serve families with children in this age range, regardless of the age of child at enrollment. To improve data quality and rigor, the preferred schedule of administering these measures is every 6 months.

Adopting this common set of indicators in both the parenting and child development domains, coupled with our Phase I recommendations, will contribute to our understanding of how different families, facing different challenges, respond to different home visiting interventions. Early home visiting programming has made great strides over the past 20 years in both defining program quality and structure. The core domains and related indicators highlighted by the DPI project reflect a growing consensus regarding the likely ways in which investments in high-quality early home visiting may alter the life trajectory of the young children, parents, and families being served. Maximizing the return from investments in more rigorous measurement approaches requires a comparable investment in staff training in how home visitors administer these measures, careful supervision and support to insure consistent practice around data collection, and maintenance of a participant-level database that allows programs to examine outcomes for specific subpopulations. Consistently measuring outcomes in multiple domains offers an important tool for understanding the collective impacts of diverse investments in diverse home visiting programs. Much remains to be learned about the extent to which these interventions achieve meaningful and lasting change across diverse participant populations and across diverse community settings. Fostering continuous learning and program improvement will require continued experimentation with different measurement strategies and exploration of other potential outcomes. As such, researchers are encouraged to augment these recommendations with additional ways of measuring impacts and sharing lessons learned with the field.

To further advance this discussion, the DPI team will engage in a number of educational and dissemination activities over the next several months. Specifically, we will present the final recommendations individually to a range of key stakeholders currently involved in implementation home visiting programs at both the state and federal levels. These stakeholders will include:

- The National Home Visiting model developers participating in the Alliance;
- The measurement committee of the Association of State and Tribal Home Visiting Initiatives (ASTHVI);
- The Maternal, Infant and Early Childhood Home Visiting (MIECHV) central office staff and Regional Project Officers;
- The Network of Infant/Toddler Researchers (NITR) Steering Committee; and
- The Home Visiting Applied Research Collaborative (HARC) or other groups funded by HRSA to advance home visiting research.

Following discussions with these stakeholder groups, we will discuss the recommendations more broadly via webinars and conference presentations.

APPENDIX A: LIST OF MEASURES EXAMINED

Parenting Measures

- Adult Adolescent Parenting Inventory (AAPI-2); Bavolek, S. & Keene, R. (2010). *The Adult-Adolescent Parenting Inventory Second Edition Handbook: Assessing high-risk parenting attitudes and behaviors*. Asheville, NC: Family Development Resources.
- Assessment of Parenting Tool; Moran, T., Troutman, B., Franklin, C. & Evenson, A. (2012). Initial validation of the Assessment of Parenting Tool: A task- and domain-level measure of parenting self-efficacy for parents of infants from birth to 24 months of age. *Infant Mental Health Journal*, 37(3), 222–234.
- Conflict Tactics Scale - Parent Child (CTS-PC); Straus, M., Hamby, S., Finkelhor, D., Moore, D., & Runyan, D. (1998). Identification of child maltreatment with the Parent-Child Conflict Tactics Scales (CTSPC): Development and psychometric data for a national sample of American parents. *Child Abuse and Neglect*, 22(4), 249–270.
- Dyadic Assessment of Naturalistic Caregiver-child Experiences (DANCE); Olds, D., Donelan-McCall, N., O'Brien, R., MacMillan, H., Jack, S., Jenkins, T., . . . Beeber, L. (2013). Improving the Nurse-Family Partnership in community practice. *Pediatrics*, 132, S110–S117.
- Early Home Learning Environment Index; Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2008). Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school. *Journal of Social Issues*, 64(1), 95–114.
- Family Crisis Oriented Personal Evaluation Scale (F-COPES); McCubbin, H. I., Olson, D., & Larsen, A. (1996). Family Crisis Oriented Personal Evaluation Scales (F-COPES). In H. I. McCubbin, A. I. Thompson, and M. A. McCubbin (Eds.), *Family assessment: Resiliency, coping, and adaptation – Inventories for research and practice* (pp. 455-508). Madison, WI: University of Wisconsin System.
- Family Map; Whiteside-Mansell, L. Bradley, R., Conners, N., & Bokony, P. (2007). The Family Map: Structured family interview to identify risks and strengths in Head Start families. *NHSA Dialog: A Research-To-Practice Journal for the Early Intervention Field*, 10(3-4), 189–209.
- Family Assessment of Strengths Tool (FAST); Slack, K. S., & Berger, L. M. (2016). *Family Assessment of Strengths Tool*. Unpublished measure.
- Healthy Families Parenting Inventory (HFPI); Krysik, J. & LeCroy, C. W. (2012). Development and initial validation of an outcome measure for home visitation: The Healthy Families Parenting Inventory. *Infant Mental Health Journal*, 33(5), 496–505.
- Home Observation Measurement of the Environment-Short Form (HOME-SF); Baker, P. C., & Mott, F. L. (1989). *National Longitudinal Survey of Youth, Children, and Young Adults Handbook*. Columbus, OH: Center for Human Resource Research.
- Infant Caregiving Inventory – Revised; Parks, P. L., & Smeriglio, V. L. (1986). Relationships among parenting knowledge, quality of stimulation in the home and infant development. *Family Relations*, 35(3) 411–416.

- Knowledge of Effective Parenting Scale (KEPS); Morawska, A., Sanders, M. R., & Winter, L. (2007). *The Knowledge of Effective Parenting Scale*. Brisbane, QLD, Australia: The School of Psychology, The University of Queensland.
- Keys to Interactive Parenting Scales (KIPS); Comfort, M. & Gordon, P. R. (2006). The Keys to Interactive Parenting Scale (KIPS): A practical observational assessment of parenting behavior. *NHSA Dialog: A Research-To-Practice Journal for the Early Intervention Field*, 9(1), 22–48.
- Knowledge of Infant Development Inventory (KIDI); MacPhee, D. (1981). *Knowledge of Infant Development Inventory*. Princeton, NJ: Educational Testing Service.
- Life Skills Progression (LSP); Wollesen, M. A. & Peifer, K. (2006). *Life Skills Progression: An outcome and intervention planning instrument for use with families at risk*. Baltimore, MD: Brookes.
- McMaster Family Assessment Device; Epstein, N. B., Baldwin, L. M., & Bishop, D. S. (1983). The McMaster Family Assessment Device. *Journal of Marital and Family Therapy*, 9(2), 171–180.
- Parent Behavior Checklist; Fox, R. A. (1994). *Parent Behavior Checklist*. Brandon, VT: ProEd.
- Parent/Caregiver Involvement Scale (P/CIS); Farran, D. C., Kasari, C., Comfort, M., & Jay, S. (1986). *Parent/Caregiver Involvement Scale*. Unpublished rating scale. Available from Dale Farran, Department of Child Development and Family Relations, University of North Carolina, Greensboro, N.C., 27412-5001.
- Parent Child Relationship Inventory; Gerard, A. B. (1994). *Parent-Child Relationship Inventory*. Los Angeles, CA: Western Psychological Services.
- Parent-Child Observation Guides (PCOG); Bernstein, V. J., Percansky, C., & Hans, S. L. (1987, April). *Screening for Social–emotional Impairment in Infants Born to Teenage Mothers*. Paper presented at the biennial meeting of the Society for Research in Child Development, Baltimore, MD.
- Parent-Infant Relationship Global Assessment Scale (PIR-GAS); Zero to Three/National Center for Clinical Infant Programs (1994). *Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood (PIRGAS)*. Arlington, VA: Zero to Three/National Center for Clinical Infant Programs.
- Parent Opinion Questionnaire; Azar, S. T., Robinson, D. R., Hekimian, E. & Twentyman, C. T. (1984). Unrealistic expectations and problem-solving ability in maltreating and comparison mothers. *Journal of Consulting and Clinical Psychology*, 52, 687–691.
- Parental Stress Scale; Berry, J. O. & Jones, W. H. (1995). The Parental Stress Scale: Initial psychometric evidence. *Journal of Social and Personal Relationships*, 12(3), 463–472.
- Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO); Roggman, L. A., Cook, G. A., Innocenti, M. S., Norman, V. J., Christiansen, K., & Anderson, S. (2013). *Parenting interactions with children: Checklist of observations linked to outcomes*. Baltimore, MD: Brookes.
- Parenting Practices Inventory; Webster-Stratton, C., Reid M. J., Hammond, M., (2015). *Parenting Practices Inventory*. Seattle, WA: The Incredible Years.
- Parenting Scale; O’Leary, S. G., Arnold, D. S., Wolff, L. S., & Acker, M. M. (1993). *Parenting Scale*. Stony Brook, NY: Psychology Department, Stony Brook University.

- Parenting Sense of Competence Scale (PSOC); Johnston, C., & Mash, E. J. (1989). A measure of parenting satisfaction and efficacy. *Journal of Clinical Child Psychology, 18*(2), 167–175.
- Parenting Stress Index - Short Form (PSI/SF); Abidin, R. R. (2012). *Parenting Stress Index - Short Form (4th Edition)*. Lutz, FL: Psychological Assessment Resources, Inc.
- Pearlin-Schooler Personal Mastery Scale; Pearlin, L. I., & Schooler, C. (1978). The structure of coping. *Journal of Health and Social Behavior, 19*, 2–21.
- Perceived Stress Scale; Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396.
- Positive Discipline Parenting Scale; Hamilton, W. K. & Carroll, P. (2016). Positive Discipline Parenting Scale: Reliability and validity of a measure. *Journal of Individual Psychology, 72*(1), 60–74.
- Postpartum Worry Scale-Revised; Moran, T. E., Polanin, J. R., & Wenzel, A. (2014). The Postpartum Worry Scale-Revised: An initial validation of a measure of postpartum worry. *Archives of Women Mental Health, 17*(1) 41–48.
- Psychological Empowerment Scale (PES); Akey, T. M., Marquis, J. G. & Ross, M. E. (2000). Validation of scores on the Psychological Empowerment Scale: a measure of empowerment for parents of children with a disability. *Educational and Psychological Measurement, 60*(3), 419–438.
- Support and Resiliency Inventory-Spouse; Bowen, G. L. & Martin, J. (2011). The 2011 Support and Resiliency Inventory (SRI): Key spouses pilot test summary report. Prepared for Flying Bridge Technologies.
- Three-Bag Assessment- Child-Parent Interaction Rating Scales; Brady-Smith, C., O’Brien, C., Berlin, L., Ware, A., & Brooks-Gunn, J. (1999). *Early Head Start Research and Evaluation Project: Child-Parent Interaction Rating Scales for the Three-Bag Assessment, 24-Month Wave*. New York, NY: National Center for Children and Families, Teachers College, Columbia University.

Child Development Measures

- Ages and Stages Questionnaire, 3rd edition – Communication domain; Squires, J., & Bricker, D. (2009). *Ages & Stages Questionnaires, Third Edition (ASQ-3)*. Baltimore, MD: Brookes Publishing.
- Assessment, Evaluation, and Programming System for Infants and Children (AEPS); Bricker, D., Pretti-Frontczak, K., Johnson, J., Straka, E., Capt, B., Slentz, K., & Waddell, M. (2004). *Assessment, Evaluation, and Programming System for Infants and Children (AEPS), Second Edition*. Baltimore, MD: Brookes.
- Battelle Developmental Inventory (BDI); Newborg, J. (2005). *Battelle Developmental Inventory, Second Edition*. Itasca, IL: Riverside Publishing.
- Behavior Assessment System for Children (BASC-3); Reynolds, C. R. & Kamphaus, R. W. (2015). *Behavior Assessment System for Children, Third Edition*. San Antonio, TX: Pearson.
- Behavior Problems Index; Peterson, J. L., & Zill, N. (1986). *Marital disruption, parent-child relationships, and behavioral problems in children*. *Journal of Marriage and the Family, 48*(2), 295–307.

- Bracken Basic Concept Scale; Bracken, B. A. (2006). *Bracken Basic Concept Scale, Third Edition*. San Antonio, TX: Pearson.
- Brief Infant Toddler Social Emotional Assessment (BITSEA); Briggs-Gowan, M. & Carter, A. S. (2006). *Brief Infant Toddler Social Emotional Assessment (BITSEA)*. San Antonio, TX: Pearson.
- Brigance Inventory of Early Development III (IED III); French B. F. (2013). *Brigance Inventory of Early Development, Third Edition (IED-III): IED III standardization and validation manual*. North Billerica, MA: Curriculum Associates.
- Carolina Curriculum for Infants and Toddlers with Special Needs (CCITSN) and Carolina Curriculum for Preschoolers with Special Needs (CCPSN); Johnson-Martin, N. M., Attermeier, S. M., & Hacker, B. J. (2004). *The Carolina Curriculum for Infants and Toddlers with Special Needs (CCITSN), Third Edition*. Baltimore, MD: Brookes.
- Denver Developmental Screening Test II; Frankenburg, W. K., Dodds, J., Archers, P., Shapiro, H., & Bresnick, B. (1990). *Denver Developmental Screening Test, Second Edition*. Denver, CO: Denver Developmental Materials.
- Developmental Assessment of Young Children, 2nd edition (DAYC-2); Voress, J. K., & Maddox, T. (2012). *Developmental Assessment of Young Children, Second Edition*. Austin, TX: Pro-Ed.
- Developmental Observation Checklist System (DOCS); Hresko, W., Miguel, S., Sherbenou, R., & Burton, S. (1994). *Developmental Observation Checklist System*. Austin, TX: Pro-Ed.
- Developmental Profile 3 (DP-3); Alpern, G. D. (2007). *Developmental Profile 3*. Los Angeles, CA: Western Psychological Services.
- Devereux Early Childhood Assessment (DECA); LeBuffe, P. A. & Naglieri, J. A. (1999). *The Devereux Early Childhood Assessment*. Lewisville, NC: Kaplan Press.
- Early Communication Indicator; Luce, G. J., Linebarger, D. L., Greenwood, C. R., Carta, J. J., Walker, D., & Leitschuh, C. (2001). Developing a general outcome measure of growth in expressive communication of infants and toddlers. *School Psychology Review, 30*(3), 383–406.
- Early Coping Inventory; Zeitlin, S., Williamson, G. G., & Szczepanski, M. (1988). *Early Coping Inventory*. Bensenville, IL: Scholastic Testing Service.
- Early Learning Accomplishment Profile (E-LAP); Hardin, B. J., & Peisner-Feinberg, E. S. (2001). *The Early Learning Accomplishment Profile*. Lewisville, NC: Kaplan Early Learning Company, Chapel Hill Training Outreach Project.
- Eyberg Child Behavior Inventory; Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory & Sutter-Eyberg Student Behavior Inventory-Revised: Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Hawaii Early Learning Profile (HELP), Birth to 3 and for Preschoolers. Vort Corporation. (2004). *Hawaii Early Learning Profile*. Palo Alto, CA: Vort Corporation.
- Infant-Toddler Developmental Assessment; Provence, S., Erikson, J., Vater, S., & Palmeri, S. (1995). *Infant-Toddler Developmental Assessment*. Chicago, IL: Riverside Publishing Company.
- Infant-Toddler Symptom Checklist; DeGangi, G., Poisson, S., Sickel, R., & Santman Wiener, A. (1999). *Infant-Toddler Symptom Checklist*. Tucson, AZ: Therapy Skill Builders.

- MacArthur-Bates Communicative Development Inventories – Short Form (CDI); Fenson, L., Pethick, S., Renda, C., Cox, J. L. Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics*, 21(1), 95–116.
- Mental Development Index scale in Bayley Scales of Infant and Toddler Development, 2nd edition; Bayley, N. (1993). *Bayley Scales of Infant Development, Second Edition*. San Antonio, TX: The Psychological Corporation.
- Parents' Evaluation of Developmental Status (PEDS); Glascoe, F. P. (1997). *Parents' Evaluations of Developmental Status: A method for detecting and addressing developmental and behavioral problems in children*. Nashville, TN: Ellsworth & Vandermeer Press.
- Preschool Language Scale (PLS-5); Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool Language Scale, Fifth Edition*. San Antonio, TX: Pearson.
- Receptive and Expressive Emerging Language Scales (REEL-3); Bzoch, K. R., League, R., & Brown, V. L. (2006). *The Receptive-Expressive Emergent Language Scale-3*. Austin, TX: Pro-Ed.
- Social Skills Improvement System (formerly Social Skills Rating System) – Social Skills Scale; Gersham, F. M. & Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales*. Minneapolis, MN: NCS Pearson.
- Transdisciplinary Play-Based Assessment (TPBA2); Linder, T. (2008). *Transdisciplinary Play-Based Assessment: A functional approach to working with young children, Second Edition*. Baltimore, MD: Brookes.
- Vineland Adaptive Behavior Scale – Vineland Social-Emotional Early Childhood Scale (Vineland SEEC); Sparrow, S. S., Cicchetti D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales, Second Edition*. Circle Pines, MN: American Guidance Service.

APPENDIX B: KEY INFORMANT INTERVIEWS

- Kim Boller, Senior Fellow, Mathematica Policy Research
- Robert Bradley, Professor, School of Social and Family Dynamics, Arizona State University
- Rachel Chazan Cohen, Associate Professor of Early Education and Care Program Director, University of Massachusetts Boston
- Brenda Jones Harden, Associate Professor, Developmental Science Program, University of Maryland
- Jon Korfmacher, Associate Professor, Erikson Institute
- Craig LeCroy, Professor, School of Social Work, Arizona State University
- Leanne Whiteside-Mansell, Professor, Department of Family and Preventive Medicine, University of Arkansas for Medical Sciences
- Nancy Donelan-McCall, Prevention Research Center for Family and Child Health, Associate Professor Pediatrics, University of Colorado Denver, Anschutz Medical Campus
- Lorraine McElvey, Associate Professor, Department of Family and Preventive Medicine, University of Arkansas for Medical Sciences
- Ron Prinz, Director of Parenting & Family Research Center, Department of Psychology, University of South Carolina
- Lauren Supplee, Director of Division of Family Strengthening, Office of Planning, Research and Evaluation
- Cheri Vogel, Senior Researcher, Mathematica Policy Research
- Yange Xue, Senior Researcher, Mathematica Policy Research
- Marty Zaslow, Director of the Office for Policy and Communications, Society for Research in Child Development

APPENDIX C: ADVISORY BOARD MEETING ATTENDEES

- David Bard, Assistant Professor, Pediatrics, University of Oklahoma Health Sciences Center
- Tiffany Burkhardt, Researcher, Chapin Hall at the University of Chicago
- Deborah Daro, Senior Research Fellow, Chapin Hall at the University of Chicago
- Kathryn Harding, Senior Director of Research, Prevent Child Abuse America (Healthy Families America)
- Janet Horras, State Home Visitation Director, Iowa State Department of Public Health
- Annette Jacobi, Chief, Family Support and Prevention Services, Oklahoma State Department of Health
- Cara Karter, Project Assistant, Chapin Hall at the University of Chicago
- Allison Kemner, Senior Director, Research and Evaluation, Parents as Teachers
- Sacha Klein, Assistant Professor, School of Social Work, Michigan State University
- Jon Korfmacher, Associate Professor, Erikson Institute
- Craig LeCroy, Professor, School of Social Work, Arizona State University
- Susan Manning, Maternal Child Health Epidemiologist, Massachusetts State Department of Public Health
- Molly O'Fallon, Director of Program Quality, Nurse Family Partnership
- Sarah Wagener, Research Assistant, Chapin Hall at the University of Chicago

APPENDIX D: RECOMMENDED INDICATORS

Parenting Capacity: Using Parent Self-Assessment

Indicator	Percent of families enrolled in home visiting who report improvement in mobilizing resources.
Operational Definition	Type of measure: Outcome.
	Population: Families enrolled in home visiting.
	Numerator: Number of families enrolled in home visiting who report improvement in mobilizing resources based on repeated assessments.
	Denominator: Number of families enrolled in home visiting with repeated assessments.
Definition of Improvement	Families report greater capacity to access needed resources from local providers (utilizing services, identify local providers, comfort in navigating service system, etc.).
Data Source	Family self-report.
Measurement Tool	Healthy Families Parenting Inventory (HFPI): Mobilizing resources subscale.
Reliability/Validity	Krysiak & LeCroy, 2012
Data Collection	Parent completes tool at baseline and follow-up point (e.g., 6-months or at termination).
Estimated Time	Less than five minutes.

Indicator	Percent of families enrolled in home visiting who report improved parent-child behaviors.
Operational Definition	Type of measure: Outcome.
	Population: Families enrolled in home visiting.
	Numerator: Number of families enrolled in home visiting who report improved parent-child behaviors based on repeated assessments.
	Denominator: Number of families enrolled in home visiting with repeated assessments.
Definition of Improvement	Families report more positive child management skills, more frequent positive interactions with child (reading, praising, calmer response to managing behaviors, etc.).
Data Source	Family self-report.
Measurement Tool	Healthy Families Parenting Inventory (HFPI): Parent-child behavior subscale.
Reliability/Validity	Krysiak & LeCroy, 2012
Data Collection	Parent completes tool at baseline and follow-up point (e.g., 6 months or at termination).
Estimated Time	Less than five minutes.

Indicator	Percent of families enrolled in home visiting who report improvement in the quality of their child rearing environment.
Operational Definition	Type of measure: Outcome.
	Population: Families enrolled in home visiting.
	Numerator: Number of families enrolled in home visiting who report improved quality in their child rearing environment based on repeated assessment.
	Denominator: Number of families enrolled in home visiting with repeated assessments.
Definition of Improvement	Families report that their child rearing environment is safer and more supportive of positive child development (appropriate limit settings, planned activities, safe, home organized, etc.).

Data Source	Family self-report.
Measurement Tool	Healthy Families Parenting Inventory (HFPI): Quality of Home Environment subscale.
Reliability/Validity	Krysiak & LeCroy, 2012
Data Collection	Parent completes tool at baseline and follow-up point (e.g., 6-months or at termination).
Estimated Time	Less than five minutes.

Parent Capacity: Using Home Visitor Assessment

Indicator	Percent of mothers enrolled in home visiting who demonstrate improved capacity to meet their child's needs.
Operational Definition	Type of measure: Outcome
	Population: Mothers enrolled in home visiting
	Numerator: Number of mothers enrolled in home visiting who demonstrate improved capacity to meet their child's needs based on repeated assessments.
	Denominator: Number of mothers enrolled in home visiting with repeated assessments.
Definition of Improvement	Home visitor assesses the mother's caregiving capacity and immediate home environment to be more conducive to positive child development.
Data Source	Home visitor assessment
Measurement Tool	HOME-SF
Reliability/Validity	Mott, 2004
Data Collection	Home visitor completes the tool at baseline and follow-up (e.g., 6 months, termination)
Estimated Time	Less than ten minutes following a home visit
Key Strengths and Limitations	Because the full HOME Inventory is currently under revision to address concerns regarding unbiased application across diverse populations, we contacted the instrument's developer (Robert Bradley) to determine when a revision might be available. Bradley has agreed to work with us in making appropriate revisions.

Parent Capacity: Home Visitor Observation

Indicator	Percent of mothers enrolled in home visiting who demonstrate more frequent parent-child interactions.
Operational Definition	Type of measure: Outcome.
	Population: Mothers enrolled in home visiting.
	Numerator: Number of mothers enrolled in home visiting who demonstrate improved capacity to meet their child's needs based on repeated assessments.
	Denominator: Number of mothers enrolled in home visiting with repeated assessments.
Definition of Improvement	Home visitor assesses the mother's caregiving capacity and immediate home environment to be more conducive to positive child development.
Data Source	Direct observation.
Measurement Tool	DANCE (NFP programs)/PICCOLO (other home visiting programs).
Reliability/Validity	PICCOLO: Roggman et al, 2013. DANCE: Donelan-McCall et al, 2012.
Data Collection	Home visitor observes parent-child interaction at baseline and repeated follow-up points.
Estimated Time	PICCOLO: 10 minutes to administer; 2 minutes to score. DANCE: 5 to 8 minutes to administer; 10 minutes to code.

Key Strengths and Limitations	Although observational tools like these have historically not been adopted by most agencies because of cost and difficulty in administration, home visiting programs are increasingly embracing these strategies and are becoming more comfortable incorporating them into their standard practice precisely because of their instrumental value in improving practice. We believe the potential strength of this strategy (as evidenced by the MIECHV recommendation that home visiting programs adopt them) provide confidence that such a tool will, in the near term, be a common element in all home visiting programs.
-------------------------------	--

Child Development: Socio-Emotional Development (For children 18-36 months)

Indicator	Percent of children enrolled in home visiting who demonstrate improved social and emotional competencies.
Operational Definition	Type of measure: Outcome.
	Population: Children enrolled in home visiting.
	Numerator: Number of children enrolled in home visiting who demonstrate improved social and emotional competencies based on repeated assessments.
	Denominator: Number of children enrolled in home visiting with repeated assessments.
Definition of Improvement	Parent reports child demonstrates improved social and emotional competencies (improved externalizing and internalizing behaviors, less dysregulation and fewer maladaptive habits).
Data Source	Parent Report.
Measurement Tool	Infant Toddler Social Emotional Assessment (ITSEA/BITSEA).
Reliability/Validity	BITSEA; Briggs-Gowan and Carter, 2005.
Data Collection	Parent completes the tool at baseline and repeated follow-up points.
Estimated Time	7 to 10 minutes
Key Strengths and Limitations	The BITSEA was normed on a sample that was not nationally representative; the sample excluded children who, at birth, were expected to have severe developmental delays and excluded parents who could not speak English. Strengths of the BITSEA include that it is available in English and Spanish and can be administered to both parents and primary caregivers. While this measure is widely used in early home visiting and other early intervention programs, BITSEA may not be highly sensitive to detecting change over time.

Child Development: Language Development (for children 18-36 months)

Indicator	Percent of children enrolled in home visiting who demonstrate enhanced language development and age-appropriate skills.
Operational Definition	Type of measure: Outcome
	Population: Children enrolled in home visiting between the ages of 18 and 36 months
	Numerator: Number of children 18 to 36 months old enrolled in home visiting who demonstrate enhanced language development and age-appropriate skills based on repeated assessments.
	Denominator: Number of children 18 to 36 months old enrolled in home visiting with repeated assessments.
Definition of Improvement	Home visitor assesses the mother's caregiving capacity and immediate home environment to be more conducive to positive language development.
Reliability/Validity	Fenson et al., 2000; Vogel et al., 2015; Bates & Goodman, 1997
Data Collection	Parent report
Measurement Tool	MacArthur-Bates Communicative Development Inventories (CDI)– Short Form
Estimated Time	10 minutes
Key Strengths and Limitations	Providers have found the CDI easy to administer and parents generally find it useful. Reliability and validity have been found to be acceptable. It is relatively low cost and well supported. While not appropriate for predicting language development under the age of one, the Level II toddler form is appropriate for assessing language development for children 16 to 30 months of age.