



**RESOURCES FOR MEASURING SERVICES AND OUTCOMES IN
HEAD START PROGRAMS SERVING INFANTS AND TODDLERS**

April 2011



U.S. Department of Health and Human Services
ADMINISTRATION FOR
CHILDREN & FAMILIES





RESOURCES FOR MEASURING SERVICES AND OUTCOMES IN HEAD START PROGRAMS SERVING INFANTS AND TODDLERS

April 2011

Prepared by:

Ellen Eliason Kisker
Kimberly Boller
Charlotte Cabili
Charles Nagatoshi
Celina Kamler
Cleopatra Jacobs Johnson
Yange Xue
Andrea Mraz Esposito
Jamila Henderson
Subuhi Asheer
Christine Sciarrino
Vinita Jethwani
Teresa Zavitsky
Melissa Ford
John M. Love

Mathematica Policy Research, Inc.

Prepared for:

Rachel Chazan Cohen

Child Outcomes Research and Evaluation
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Services

Summary of 2011 Updates

The purpose of this resource document is to provide information about existing screening and assessment instruments designed for use with children under age 3 and their families, as well as instruments designed for assessing services provided by programs serving them. It is a living document to which information on new instruments is added so that new tools and approaches can be shared. The first update to this resource document occurred in 2011.

The update consists of 29 updated and newly added instruments. In particular, we updated 16 current profiles of instruments for which the publisher had made newer editions available by the 2010 selection period. We added 13 instruments that were used in the Early Head Start Family and Child Experiences Survey (Baby FACES) and were more widely used among Early Head Start programs in recent years.

Substantive updates are located in Appendix C and Section 3. Appendix C tabulates instrument characteristics and provides links to the updated and newly added individual instrument profiles. In Section 3, which summarizes the information described for each instrument, we have added text to specify how descriptions differ in the new and updated instruments. The main changes to the descriptive approach include richer, lengthier descriptions of the instrument. For example:

- The **Publisher** section of the profile table includes a link to the instrument on the publisher's website, when available.
- The instrument **Description** includes information about the stimuli and manipulatives used to aid administration. The Description also documents changes from the previous version of the instrument.
- A section was added, **Other Languages**, to describe translations of the instrument. When available, details on the norming sample, reliability, validity, and English language equivalence or comparability are provided.
- **Content validity** was expanded to include information related to literature reviews and factor analysis of items.
- A section was added, **Bias Analysis**, to describe tests for whether instruments perform differently among subgroups (e.g., age, race/ethnicity, or spoken language) when differences are not expected.

Abstract

This document contains resources to help Head Start programs that serve pregnant women and families with infants and toddlers develop a performance measurement plan and carry out data collection that will support their continuous program improvement efforts. These performance measures activities should build upon existing screening and assessment activities required by the Head Start Program Performance Standards. This document discusses the importance and development of a comprehensive plan and presents profiles of instruments that may be useful to programs. Because we anticipate that it will be used under different circumstances for different purposes, we present the background information in a book format, and the entire document on a compact disk, to enable users to search for the sections and measures that apply to them. This format will also support the continued evolution of the document, which is intended to be a “living” document to which information on new instruments can be added, through which new tools and approaches can be shared, and in which other resources that individual programs find useful can be compiled. These materials will be most useful when used in consultation with an assessment expert.

THE INCLUSION OF AN INSTRUMENT IN THIS RESOURCE DOCUMENT DOES NOT CONSTITUTE ENDORSEMENT OF THE INSTRUMENT BY THE AUTHORS, MATHEMATICA POLICY RESEARCH, OR THE U.S. GOVERNMENT.

Table of

Contents

1. PERFORMANCE MEASUREMENT	1
2. FORMULATING A PERFORMANCE MEASUREMENT PLAN	7
3. INFORMATION INCLUDED FOR EACH INSTRUMENT	21
REFERENCES	35
APPENDIX A:	39
WORKSHEET TO HELP DEVELOP A COMPREHENSIVE PLAN FOR GATHERING & ANALYZING DATA	
APPENDIX B:	43
PERFORMANCE MEASURES FRAMEWORK FOR HEAD START PROGRAMS SERVING INFANTS AND TODDLERS	
APPENDIX C:	51
LIST OF MEASURES USED IN THE NATIONAL EARLY HEAD START RESEARCH AND EVALUATION PROJECT	

PERFORMANCE MEASUREMENT

Section 1



Performance measurement includes data collection and aggregation activities that give staff members the opportunity to look at how their program is doing, that is, to determine whether they are providing the services they intend to provide and how children and families are faring. Program staff can use this information in planning for continuous program improvement activities, and it can be shared with stakeholders such as parents and funders.

Head Start and later Early Head Start, as national laboratories for early childhood education, have been leaders in developing outcomes-oriented accountability. It began developing performance measures in 1995, partly to respond to the Government Performance and Results Act (GPRA) requirements. In 1997, the Office of Head Start launched the Family and Child Experiences Survey (FACES) to collect data on child and family outcomes, as well as program services and management systems, for a large

nationally representative sample of children and families in Head Start programs (Administration on Children, Youth and Families 2001b). This study links the development of children and families with their experiences in Head Start. Following the reauthorization of Head Start in 1998, Head Start programs were required to include child outcomes in their self-assessment process by 2003.

In 2007, the Office of Head Start launched the Early Head Start Family and Child Experiences Survey (Baby FACES) to collect data on child and family outcomes, as well as program services and management systems, for a large nationally representative sample of Early Head Start programs and of enrolled newborn and 1-year-old children. This study, currently underway, links the development of children and families with their experiences in Early Head Start. Although Early Head Start programs have not been required to report child outcomes, many have started to try to define and measure outcomes, for several reasons. Some programs are doing so in conjunction with performance measurement in Head Start because they operate within Head Start programs. Some are responding to other funders' requirements. Finally, some simply want to improve their services to families with infants and toddlers. In Spring 2001, the Early Head Start Technical Work Group recommended that the Office of Head Start move forward to develop performance measures for Head Start programs serving infants and toddlers (Early Head Start

and Migrant Head Start programs) to support programs' efforts.

The development of performance measures for Head Start programs serving pregnant women and families with infants and toddlers has two purposes. These are (1) to create performance measures sensitive to the developmental stages from pregnancy to age 3 that can be used at both national and local levels for learning about child and family experiences along with program services and management systems; and (2) to provide guidance to local programs in defining and measuring outcomes and using this information for continuous program improvement.

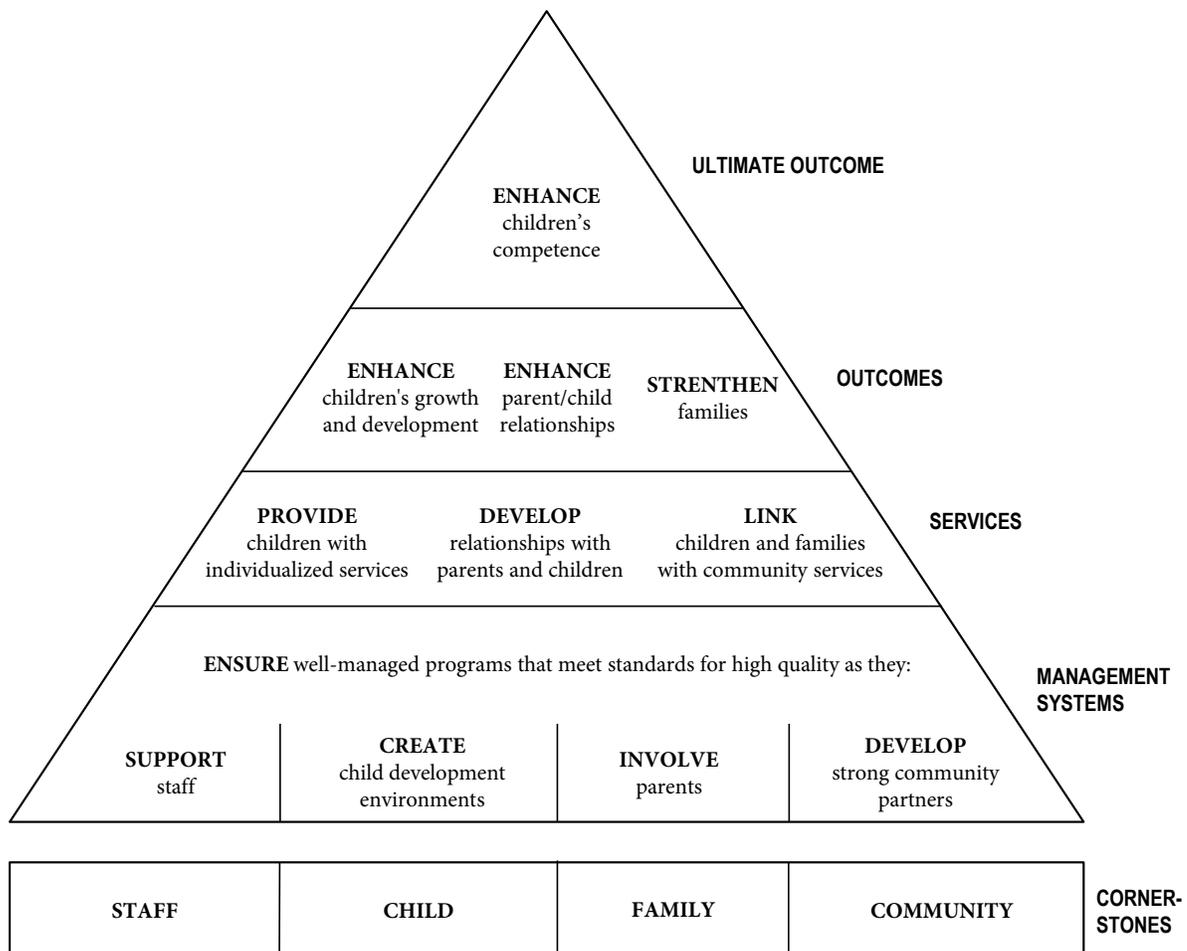
To fulfill these purposes, the Head Start pyramid, representing the Head Start performance measures framework, was expanded to reflect the unique features of programs serving infants and toddlers (see Figure 1 and Appendix B). The ultimate goal of Head Start programs that serve pregnant women and families with infants and toddlers is the same as that of Head Start programs that serve preschool children—enhancing children's competence.¹ The main outcomes supporting this ultimate goal—the blocks in the pyramid—have been expanded to reflect the central role of

¹ The Head Start goal, traditionally stated as children's "social competence," was shortened to "competence" because the interpretation of social competence in the Head Start pyramid has sometimes been narrower than was intended. Competence is the child's everyday effectiveness in dealing with his or her present environment and later responsibilities in school and life.

relationships in supporting attainment of this goal in families with younger children (U.S. Department of Health and Human Services [DHHS] 1994). Enhancing children’s growth and development, enhancing parent-child relationships, and strengthening families as primary nurturers of their children are outcomes that support the ultimate goal of enhancing children’s competence. Supporting these outcomes are objectives related to program services, including the objectives of providing children with individualized services, developing relationships with parents and children, and

linking children and families to needed community services. The objective of ensuring well-managed programs that meet standards for high quality in supporting staff, providing early childhood environments, involving parents, and developing community partnerships serves as the foundation of the pyramid. The pyramid rests on the four cornerstones recommended by the Advisory Committee on Services for Families with Infants and Toddlers—staff development, child development, family development, and community building (DHHS 1994).

Figure 1 CONCEPTUAL FRAMEWORK FOR PROGRAMS SERVING INFANTS & TODDLERS





The national Early Head Start Research and Evaluation Project (EHSREP) studied early program implementation and impacts on children and families (see Box 1). From this study, we learned that each program operates with its own theory of change—that is, a theory explaining how the services they provide will improve the child and family outcomes they are focusing on to meet the needs of children and families in their community. Although all programs must implement all aspects of the pyramid, the ways that programs configure their management systems and program services (the bottom levels of the pyramid) vary widely. The specific child and family outcomes they focus on (the upper levels of the pyramid) also vary. As the program staff learns from continuous program improvement activities or as families' needs change, programs' theories of change may evolve. The evaluation showed that in general, patterns of program impacts reflected differences in theories of change. Home-based programs, which emphasized improving parenting and the home environment as an important path to improved child outcomes, had significant favorable impacts on parenting and child outcomes. Center-based programs, which emphasized direct services to children to improve outcomes, had favorable impacts on children and fewer significant impacts on parenting. The study also showed that programs that fully implemented key elements of the Head Start Program Performance Standards (elements related to all levels of the pyramid) had more

favorable impacts on a wide range of outcomes than programs that were incompletely implemented (Administration for Children and Families 2002).

A **theory of change** is a belief or set of beliefs about how program services and other factors produce changes in the desired outcomes. The theory of change is the basis for designing and implementing program services. It may be explicit in program documents or staff discussions, or it may be implicit in the decisions of program designers and implementers.

The pyramid provides a framework to help programs identify their own theory of change. Together, the pyramid and theory of change provide guidance for examining links between program services and outcomes to inform continuous program improvement activities.

The performance measures framework is general and does not identify specific program services and outcomes that should be measured. Individual programs must identify specific services and outcomes based on their own theory of change and select appropriate instruments to fit their available resources.

We do not recommend that programs collect data related to every element of the pyramid (at least to start with). We recommend that programs collect data related to each level of the pyramid (management systems, services, and outcomes), because having such information allows staff to link information on program characteristics and services with outcomes for

children and families and to learn about both how well children and families are doing and how services might be improved to promote better outcomes.

Selecting appropriate measures is an important and complex process. This document is intended to be a resource for programs undertaking this process. It includes both screening instruments and assessment instruments that may be useful to programs as they explore how to approach performance measurement. We attempted to identify a wide range of instruments and select those likely to be most useful to Head Start programs serving pregnant women and families with infants and toddlers, but there are useful tools that are not included in this review. For example, some kinds of assessment that are useful for individualizing services, such as portfolio assessment, are not included (Martin 1999). In addition, new measures are constantly emerging, and very

recent measures may not be found here. Some programs may wish to use this resource document in consultation with an expert on screening and assessment to help them develop a plan and select instruments. (For more explanation of how we selected instruments for this resource document, see page 22.)

This resource document will be most useful to programs if it is used in conjunction with a comprehensive plan for performance measurement. In the next section, we discuss the importance of developing a plan and the elements that make up such a plan. In the final section, we describe how instruments were selected for this resource document, explain the information provided for each instrument, and present profiles of instruments that may be useful to Head Start programs that serve pregnant women and families with infants and toddlers.

Box 1

MAKING A DIFFERENCE IN THE LIVES OF INFANTS AND TODDLERS AND THEIR FAMILIES: THE IMPACTS OF EARLY HEAD START

A rigorous evaluation of Early Head Start services in 17 programs selected from the first groups of programs funded showed they had significant favorable impacts on a wide range of parent and child outcomes, some with implications for children's later school success. Findings from the study (*Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start*), using data gathered when children were age 3 and had completed the program, show that the programs sustained and broadened the pattern of impacts reported when children were age 2 (*Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families*, 2001). All Early Head Start evaluation reports are available online at [<http://www.acf.hhs.gov/programs/opre>].

Early Head Start Improved Outcomes. The national evaluation conducted by Mathematica Policy Research, Inc. and Columbia University's Center for Children and Families at Teachers College, in collaboration with the Early Head Start Research Consortium, reported that 3-year-old Early Head Start children performed significantly better on measures of cognitive, language, and social-emotional development than a randomly assigned control group. While children who participated in Early Head Start performed better than their peers who did not receive Early Head Start on all aspects of development that were assessed, both groups lagged behind 2- and 3-year-old children nationally. This may, in part, be due to the fact that Early Head Start programs must reserve at least 10 percent of their slots for children with disabilities, including those with developmental delays who score at the lower end of the distribution. Parents in Early Head Start scored significantly better than control group parents on measures of many aspects of the home

environment and parenting behavior. Furthermore, Early Head Start programs enhanced parents' progress toward self-sufficiency years. Early Head Start fathers benefited as well. Although the overall impacts of Early Head Start were generally modest, the pattern of favorable findings across outcomes in a wide range of key domains is promising.

Full Implementation Was Important. The impacts on children and parents were consistent with the substantial difference the programs made in families' receipt of services. Early Head Start families were significantly more likely than control families to receive a wide variety of services, but especially to receive intensive services, and to receive intensive services that focused on child development and parenting.

Implementing the Head Start Performance Standards early and well is important for maximizing impacts on children and families. The research programs were systematically rated according to the extent to which they implemented key elements of the Performance Standards. Those that fully implemented the standards demonstrated a broader pattern of significant impacts than did the programs that did not reach full implementation during the evaluation period.

Patterns of Impacts Were Consistent with Theories of Change. Analyses that compared the contribution of impacts on parenting when children were age 2 to impacts on child outcomes at age 3 in programs providing home-based or a mix of home- and center-based services generally provided support for the theories of change that staff in those programs described, with some of the effects of the program on 3-year-old children being associated with the effects on parenting when they were age 2.

FORMULATING A PERFORMANCE MEASUREMENT PLAN

Section 2



The Head Start Program Performance Standards require programs to conduct screening and assessment activities and emphasize their importance for individualizing services and informing continuous program improvement. The performance standards allow considerable flexibility in how programs meet the requirements. In response to the requirements in the performance standards, Head Start programs serving pregnant women and families with infants and toddlers are already collecting data. Head Start programs serving pregnant women and families with infants and toddlers may use screening and assessment instruments for different purposes, including:

- ***To Support Development and Learning.*** Child assessment results can tell caregivers and teachers what each child can do and what he or she is ready to learn next. Family or parent assessment results can help program staff identify family or parent strengths, needs, and concerns and tailor services to the family. Over time, assessment

results can demonstrate how each child and family is progressing. Assessment results can also help staff communicate with family members about their children's needs and progress, as well as their own needs and progress. This information can help staff individualize services and improve them over time.

- ***To Identify Special Needs and Concerns.***

Because of the cost of in-depth assessments, screening is usually the first step in identifying special needs. Children or families for whom an in-depth assessment is indicated are often referred to a physician or other expert for a complete evaluation. Screening results provide the information needed for referrals to other agencies, such as Part C agencies, to obtain services for children with disabilities.

- ***To Evaluate the Program and Monitor Trends.***

For this purpose, child and family screening and assessment data may be aggregated and used to inform continuous program improvement efforts. Program assessment data and feedback from Office of Head Start monitoring may also be used for this purpose. Aggregated screening and assessment results can inform staff about how well the program is meeting child, family, and community needs.

Screening and assessment results for individual children, along with other information from parents and caregivers, are

needed to help staff tailor services for those children. When the same instruments are used for all children in a program, aggregating data across families can provide a picture of how children and families in the program are doing overall. When this aggregate information on child and family outcomes is linked to information on services and other program characteristics, it can provide insights that are useful to staff members in their continuous program improvement efforts. Aggregating data provides a picture of how children and families and the program are doing (that is, it measures the program's performance). Over time, aggregated data can be used to track changes in child and family functioning, which along with information about changes in program services or characteristics, can help staff learn about what works best for particular types of children and families. The aggregated information on how children and families in the program are doing can also be useful for meeting other funders' reporting requirements.

Screening. Screening is a generic term referring to activities designed to identify individuals who have a high probability of exhibiting delayed, abnormal, or problematic development. The screening is intended to identify problems at an early stage and identify individuals for whom further, in-depth assessment activities are needed.

Assessment. Assessment is a generic term referring to procedures for obtaining systematic information on a child's, parent's, family's, or

program's strengths or needs. As noted in Chapter I, the Head Start Program Performance Standards focus on the child and family assessment purposes of identifying "(i) the child's unique strengths and needs and the services appropriate to meet those needs; and (ii) the resources, priorities, and concerns of the family and the supports and services necessary to enhance the family's capacity to meet the developmental needs of their child."

[See the Early Head Start National Resource Center's Technical Assistance Paper # 4 for more detail on screening and assessment activities.]

Performance measurement includes data collection and aggregation activities that give staff members the opportunity to look at how their program is doing, that is, to determine whether they are providing the services they intend to provide and to learn how children and families are faring. Program staff can use this information in planning for continuous program improvement activities, and it can be shared with stakeholders such as parents and funders.

Screening results for individual children and families are useful for deciding whether further assessment is needed. Screening results can also be aggregated to provide information on the extent of potential problems in the population and the need for in-depth assessments among children and families overall. Aggregated screening results and information on referrals can inform a program as to whether these in-depth assessments are happening if the

program's tracking system does not provide this information.

Assessment results for individual children can be used for planning services; in addition, they can often be aggregated to provide broader information on child outcomes. Some approaches to assessment are valuable for individualizing services, but cannot be aggregated unless they are translated into another form. For example, portfolio assessment can be extremely useful for individualizing services for children, but unless a systematic way of coding the information is developed, the results cannot be aggregated.

Programs face both a great opportunity and a significant challenge as they consider performance measurement. They have an opportunity to select instruments and collect data that best meet the needs of their families and their program, and they face the challenge of figuring out just what those instruments should be, how they should be administered, and how the information collected using those instruments should be analyzed. Meeting that challenge effectively requires programs to develop a plan that considers the purpose(s) for which the data will be used, what data are already being collected and additional data that need to be collected, and how data will be aggregated and analyzed for continuous program improvement. Taking the time to develop a comprehensive plan will help ensure that the program's resources for the required ongoing screening and assessment of individual



children and families, as well as the new performance measurement and data aggregation activities, are used well and that the activities are as useful as possible.

To formulate a comprehensive plan, programs need a clear understanding of why performance measurement is needed. What does the staff need to know to determine whether the program is achieving its intended outcomes and to individualize or optimize services for children and families? What information is essential? What, in addition, would be good to know? Some questions programs might consider to help reach this understanding are:

- What are the program's intended outcomes? What is the program trying to accomplish?
- How will the program's services influence these outcomes? What will the program do or provide for its families to achieve the desired outcomes?
- What information does the program need for individual service planning? Is this information needed at enrollment? How often and at what times during the program is this information needed? What information is already being collected (such as HSFIS data, child and family screening/assessments) that can be used for this purpose?
- What information does the program need for program planning? How often is this information needed? What information is

already being collected (such as HSFIS data, PIR data, program self-assessment information) that can be used for this purpose?

- How will the program know that staff are doing what they think they are doing?
- How will the program know how well children and families are progressing?
- Overall, are services implemented well? How will the program know services are being implemented well?
- Is the program influencing the targeted outcomes across all families served? Are particular types of families benefiting more (or less)?
- What information will help demonstrate to funders that the program is effective?
- How will the program use the information gathered for program improvement?

To make performance measurement activities as useful as possible, programs need a theory of change. This model, or theory of change, provides a framework for guiding the selection of individual instruments and for integrating information obtained from the selected instruments and other sources of information about a child and family. Such a model or theory of change simply specifies explicitly what child and family development outcomes the program is trying to improve and how the program's services will influence those outcomes (comprehensive answers to the first two questions in the list above constitute a

theory of change).² With such a model or theory of change, the staff can select instruments that will focus on measuring targeted outcomes and the key services designed to improve them. The program's theory of change can also guide the interpretation of data at both the individual and program levels and decisions about what to do in response to the information.

The performance measures pyramid, along with a theory of change, provides a framework for developing a program-specific comprehensive plan for performance measurement. We do not recommend that programs collect data related to every element of the pyramid (at least to start with). We recommend that programs collect some information related to each level of the pyramid (management systems, services, outcomes, and the ultimate outcome), because having such information allows staff to link information on program characteristics and services with outcomes for children and families and learn about both how well children and families are doing and how services might be improved to promote better outcomes. All plans should include measurement of child outcomes.

The following example illustrates how the performance measures framework and an explicit theory of change can help program staff

develop a comprehensive plan for gathering and analyzing information. A program that provides home-based services may have a theory of change that indicates that it is trying to improve child development outcomes indirectly by improving parenting and parent-child relationships. To accomplish this, it provides home visits in which staff members work with parents on specific parenting skills. The program may emphasize improving children's language development and train home visitors to work with parents on activities they can do with their child to promote learning and language development, such as reading to them regularly and providing a home environment that supports learning and language development. Based on its theory of change and emphasis on language development, this program might give priority to selecting instruments that measure children's language development (outcome), the frequency of parents' reading to their child (outcome), support for language and learning in the home environment (outcome), the frequency and quality of home visitors' interactions with families focused on children's language development (services), and supervisory practices to support home visitors' interactions with families (management system). Ideally, the data obtained with these instruments will indicate how children's language development is progressing, provide insights into how parents' progress in parenting skills may be promoting their children's language development, illuminate how program services may be contributing to parents' progress (and indirectly

² For an illustration of the application of the theory-of-change approach to identifying expected outcomes in Early Head Start, see the first report on program implementation of the national Early Head Start Research and Evaluation Project (EHSRE) (ACYF 1999, Chapter II).

to children's language development), and indicate how home visitor supervision is supporting home visitors' activities related to language development. The data may also provide insights into ways that home visitors' activities with parents can be improved to enhance children's language development further.

A comprehensive plan for gathering and using data includes several important elements.

The following questions indicate important elements that a comprehensive plan should include:

- Does the plan include clearly stated purposes for gathering and using data?
 - What instruments will be used to gather information for individual service planning and aggregate performance measurement?
 - With which children or families will each instrument be used?
 - When and how often will each instrument be administered?
 - Who will administer each instrument, and what training will they receive?
 - How will the results of administering each instrument be recorded?
 - How will the results of each instrument be kept confidential?
 - How will the results be shared with parents?
 - How will results be used to plan services for individual children and families? Will they be used for referrals to other community service providers, such as Part C agencies?
- Will the individual results be aggregated across children and families and analyzed for program planning? If so, how?
 - How will the results feed into continuous program improvement?
 - How will the results be reported to other stakeholders?
 - How will the plan be modified as you gain experience with it?

Appendix A contains a worksheet that provides one way to summarize some of this information and that may be helpful as a component of a comprehensive plan.

In developing a plan that is appropriate and feasible for its program, the staff needs to consider its priorities among information needs.

Programs may not have the resources to gather and analyze all useful data, and staff members may not have experience with aggregating results from screening and assessment instruments. Although it is important for programs to create a comprehensive plan, it may be necessary, for guiding program activities, to identify priorities that are feasible with current resources and staff experience. As staff gain more experience with collecting and aggregating data, additional elements of the plan can be implemented (if resources are available).

The kinds of resources that should be considered in developing and implementing a comprehensive plan for gathering and analyzing information include staff resources, technical resources, and financial resources.



Staff resources include the time that could be devoted to administering instruments and analyzing the results, as well as activities to build skills and obtain training. Technical resources include training and technical assistance, consultation, and computer hardware and software—as well as qualified personnel—to manage and analyze data. Financial resources include money to purchase needed materials, pay for additional training and technical assistance, and purchase computer hardware or software.

In determining measurement priorities, programs should consider giving priority to instruments that together represent a balance across elements of the pyramid for Head Start programs serving pregnant women and families with infants and toddlers. As noted earlier, lower entries in the pyramid (program management and services) support the outcomes above them. Thus, information on program management and services along with information on child and family outcomes will provide greater insights into ways in which program services can be improved to promote better outcomes at both the individual and the program levels. First priority, however, should be given to measuring child outcomes.

A comprehensive plan for gathering and analyzing data will be useful only if the staff is willing and able to implement it. In developing such a plan, program managers need to build support among staff members and prepare them to use the selected instruments. Involving key

staff members who will implement the plan in its development may promote their “buy-in” to the new activities. Involving staff will also help to illuminate the questions that the program most needs to answer. In addition, providing enough training and time to administer the selected instruments may also help staff members embrace the new activities. Finally, involving staff members in interpreting and using the results will help them see how they are useful, motivate them to administer the instruments well, and enable them to help identify needed modifications to the plan.

A key part of the development of a comprehensive plan is the selection of specific instruments that will be used. The checklist in Box 2 lists some questions to help the program staff identify instruments that will best meet its needs. In addition to these questions for considering individual instruments, it is important that, together, the selected instruments draw on multiple sources of information. To be useful for aggregating across families, they also need to be administered to all children or families (or the group of children or families for whom aggregated information is needed) in a consistent manner.

In developing a comprehensive plan, care should be taken to prevent misuses of instruments. It is not appropriate to select a subset of items from an instrument, combine items from multiple instruments, or change the wording or response categories for items in an

instrument, because the abbreviated or changed instrument may not be reliable or valid. Some instruments, however, include official subscales or subtests that may be used alone. Only if the directions for using an instrument indicate that using just a sub-scale or subtest is appropriate should staff select and use parts of instruments. To prevent misuse, it is also important that staff members who will be administering an instrument and interpreting the results have sufficient knowledge and training to enable them to do so accurately and appropriately. Lack of knowledge and understanding of an instrument can lead to its misuse.

While implementing a plan for gathering and analyzing data, a program may see the need for changes to the plan. The instruments selected initially may not work well (for example, requirements for administering them may be too difficult to meet or scoring may be too difficult), and different instruments may meet program needs better. The staff may also find that instruments selected initially do not provide all the information needed and that alternative or additional instruments may better meet the program's needs.

The development and implementation of a plan for gathering and analyzing data takes time. Figure 2 illustrates the development of a continuous improvement model in the Clayton Family Futures program in Denver, Colorado. It summarizes the steps that the program has taken to develop its model, the resources required, the timeframe for each step, and the implications of each step for the program. Over time, the program's continuous improvement activities have grown as the staff has experienced the value of the information and asked more questions about program services and how children and families are doing.

A plan for measuring outcomes can also be implemented at a broader level. Box 3 illustrates a statewide plan for collecting and analyzing data on a common set of performance measures. State Early Head Start Programs in Kansas have agreed to collect data on a common set of measures to support outcomes-based community planning.

Box 2

CHECKLIST TO HELP IDENTIFY APPROPRIATE SCREENING AND ASSESSMENT INSTRUMENTS

Below are elements of screening and assessment instruments and their use that contribute to their usefulness for Head Start programs serving pregnant women and families with infants and toddlers. For each instrument under consideration, check the box beside each element that applies. The more boxes that are checked, the better is the match between the program's needs and the instrument.

	Instruments					
The instrument measures what the program wants to know.						
The instrument was designed for the purpose for which it will be used.						
The instrument is appropriate for the cultural backgrounds of children/families who will be assessed.						
The reliability and validity of the instrument are sufficiently high for the purposes for which it will be used.						
Sufficient resources are available to obtain and use the instrument.						
It is feasible to administer the instrument according to the instrument developer's directions.						
The instrument facilitates sharing information about children (or families) with staff and parents.						
Staff members who will administer the instrument have (or will receive) the training needed to administer and score the instrument correctly.						
The instrument is appropriate for children with disabilities (or their parents).						
It is feasible to administer the instrument in settings children (or families) are comfortable with.						

Figure 2

THE DEVELOPMENT OF A CONTINUOUS IMPROVEMENT MODEL CLAYTON FAMILY FUTURES EARLY HEAD START PROGRAM

<p>Frequently Asked Questions</p> <p>Will what works in one program system work in other areas?</p> <p>Ex. The system for providing Diet and Nutrition screenings was identified as very efficient. Could we apply this system to other program areas needing a similar process?</p> <p>What actions can we take in areas that are not reaching the benchmarks established?</p> <p>Several internal actions were taken to address the challenges in providing dental screenings. However, they did not produce the desired outcomes. In response, the health team created a strong community collaboration with the School of Dentistry that was extremely effective in connecting children to dental services.</p>	<p>Step One: Setting up tracking systems and monthly reporting formats for a limited number of outcomes (a good starting point is to ask, “Are we meeting the Performance Standards in all areas?”)</p>			
	<p>Actions</p> <ul style="list-style-type: none"> Identify what you want information about Identify what is currently being tracked in those areas Identify the data tracking forms that exist and/or that need to be created Set up databases to support the collection of the information Set up monthly summary reporting formats and ongoing deadlines for the report distribution 	<p>Resources</p> <ul style="list-style-type: none"> Analyze the budget—begin to allocate funds for supporting continuous improvement implementation (start small) Designate existing and/or new staff to carry out the actions 	<p>Time Frame</p> <p>2-3 months</p>	<p>Program Implications</p> <ul style="list-style-type: none"> Questions about the link between reports and job performance Training for supervisors to effectively utilize the reporting in supervision Creation of meaningful dialogue about barriers encountered by staff in delivering quality services
<p>Developed by Chris Sciarrino, The Clayton Foundation, Denver, Colorado</p> <p>Consultation by Charmaine Lewis, Clayton Family Futures Early Head Start, Denver, Colorado</p>	<p>Step Two: Beginning to use inquiry and analysis as a method of self-evaluation, reflection, and program improvement</p>			
	<p>Actions</p> <ul style="list-style-type: none"> Disseminating summary reports to appropriate staff Program leadership leads the way for team analysis by working with the creator of the reports to identify trends, issues, and strengths. Monthly team analysis meetings are established with key staff responsible for supervision of program implementation areas. Action plans are created with time lines to address identified areas of need and how reports will be used in supervision Establishment of benchmarks for every outcome area 	<p>Resources</p> <ul style="list-style-type: none"> This model utilized 1 FTE for supporting the development and ongoing implementation of the design One formal meeting for the director to discuss the reports with the creator of the reports was established A monthly administrative team meeting focused on analyzing the reports was established Several informal discussions regarding implementation, analysis, and supervision strategies were needed to modify and adapt the approach 	<p>Time Frame</p> <p>1 year</p>	<p>Program Implications</p> <ul style="list-style-type: none"> Working with program coordinators to see reports as objective and representative of what families and children are receiving versus “what I have not done.” Beginning to use a reflective process to create change Ability to use tangible evidence is an opportunity to acknowledge and address needs Creating pride in accomplishment of benchmarks

Figure 2

Frequently Asked Questions	Step Three: Using Continuous Improvement data for program planning and communication with stakeholders			
	Actions	Resources	Time Frame	Program Implications
<p>Is it all about the numbers?</p> <p>Staff struggled with putting quality into numbers. Herein lies one of the basic challenges of the approach—how is it done so it is not seen as an either/or paradigm, either quality or accountability? It is a tendency to see these as opposing views. It is our belief that these are not separate concepts. Being able to provide consistent services (accountability) is basic to the quality and integrity of the program. It is the analysis of the data and how leadership is able to interpret its impact on quality that brings the process to life.</p>	<ul style="list-style-type: none"> Utilize reports to draw conclusions and ask questions about the EHS experience for children and families. Link this information to reflective supervision with staff. Reports collected over time are compared to demonstrate trends, highlight issues, program strengths and needs Information is used in conjunction with yearly self-assessment and community needs assessment for program planning Reports continue to be refined as adaptations needed present themselves. <ul style="list-style-type: none"> Reports are summarized quarterly for dissemination to stakeholders 	<ul style="list-style-type: none"> Meeting time Staff commitment to a set of clearly articulated and shared values- <ul style="list-style-type: none"> Quality-commitment to striving for excellence; doing the best possible job working toward a common vision Accountability-commitment to a set of clear, well defined and high standards (i.e., performance standards); demonstrating through action our ability to meet those standards Openness and collaboration-commitment to and examination of diverse perspectives and engaging in group processes and partnerships that help determine the path to high quality programs Reflection- commitment to reviewing and dialoguing about current practices along with an acceptance of one’s personal responsibility in achieving high quality Self-growth-dedication to each individual’s growth Follow through-commitment to making happen what is planned, expected, and desired 	<p>ongoing</p>	<ul style="list-style-type: none"> As coordinators become invested in and comfortable with the reports, they begin meeting together to ensure integration of Continuous Improvement efforts and to identify barriers to quality implementation of services. Analyses discussions became more complex as staff began to ask more questions. Having aggregated data readily available on a monthly basis allows for staff to easily answer 1.) Are we doing what we say we are doing? 2.) Are we accomplishing the outcomes we want to accomplish? Creating a meaningful dialogue and sharing of outcomes information with governing boards, policy council, parents. Readily available outcome data allows for timely and accurate reports to funders
<p>Developed by Chris Sciarrino, The Clayton Foundation, Denver, Colorado</p> <p>Consultation by Charmaine Lewis, Clayton Family Futures Early Head Start, Denver, Colorado</p>				

Comments from staff:

“Having information about outcomes reported to me has really helped me plan for my whole class and individual children.”

“At first it felt vulnerable and scary that my job was out there in numbers, but now I realize how helpful it is to know just what is happening and how the reports help me to offer better services.”

“We know we are ‘walking our talk’.”

“Rather than responding to a checklist for someone else, we are using the information with staff and for planning.”

Box 3

AN EXAMPLE OF A STATEWIDE APPROACH TO MEASURING OUTCOMES FOR EARLY HEAD START

Efforts to measure and report outcomes can be implemented at broader levels. In Kansas, for example, state Early Head Start programs developed and agreed to collect data on a core set of outcomes as part of a statewide system for assessing services for children and families. This system, called Connect Kansas, supports outcome-based community planning and community capacity building to create and sustain environments in which all Kansas children are safe, connected, nurtured, and supported by caring and involved adults and communities. Outcomes for Early Head Start programs were developed through four focus group discussions. These focus groups included a wide range of stakeholders, state administrators, federal Administration for Children and Families Region VII staff, Head Start Quality Improvement Center staff, Early Head Start directors, and parents.

The following core outcomes will be measured consistently by 13 state Early Head Start grantees in 32 counties. Families must be enrolled in Early Head Start for a minimum of 6 months to be included in outcomes measures.

Outcome 1: Pregnant women and newborns thrive.

- ___ % of pregnant women who receive prenatal care within the first 45 days of enrollment
- ___ % of new mothers who deliver an infant weighing 5.5 pounds or greater

Outcome 2: Infants and children thrive.

- ___ % of teachers working on or having a minimum Child Development Associate (CDA) certificate Show Breakdown:
 - ___ are working on a CDA (any stage but not yet credentialed)
 - ___ have acquired a CDA
 - ___ have an AA/AS in ECE or related field
 - ___ have a BA/BS in ECE or related field
 - ___ have a MA/MS in ECE or related field
 - ___ have other degree, specify

- ___ % of Early Head Start learning environments with a score of 5 or higher using the Thelma Harms Rating Scale (measured at entry, 6 months, 1 year, and every year thereafter. Data should be taken from the last score.)
- ___ % of Early Head Start children who are up-to-date on immunizations
- ___ % of Early Head Start children who are up-to-date on well child checks/Kan Be Healthy

Outcome 3: Children live in stable and supported families

- ___ % of parents who demonstrate improved parenting skills (measured by the Parents as Teachers Parent Knowledge Questionnaire upon entry, 6 months, 1 year, and every year thereafter)
- ___ % of enrolled families with one or more parents employed, enrolled in school, or attending a job training program 9 out of 12 months enrolled in Early Head Start Show Breakdown:
 - ___ less than 30 hours of employment
 - ___ greater than or equal to 30 hours of employment
 - ___ enrolled in school (part time or full time)
 - ___ attending a job training program
- ___ % of families who have a supportive home environment for their child with a variety of learning experiences and materials (measured by the HOME upon entry, 6 months, 1 year, and every year thereafter. Data would be taken from the last score, middle half or higher. Data will not be collected on first time pregnant women until after the birth of the baby.)

Box 3

Outcome 4: Children enter school ready to learn.

___ % of children without a diagnosed disability who demonstrate age-appropriate development in the three domains of: Intellectual, Social-Emotional, and Motor Skills (measured by the Parents as Teachers Developmental Milestone Checklist)

___ % of children who demonstrate age-appropriate language (as measured by the Early Communication Indicator, Juniper Gardens)

Other data needed for collection purposes only:

1. Total # of children enrolled in EHS who are receiving child care services.
2. Total # of non EHS children receiving quality child care services in EHS child care partnerships.
3. 3% of children identified through screening for further intervention services.
4. % of children with an IFSP, Individual Family Service Plan or IEP, Individual Education Plan, in place (IFSP/IEP denotes special services).

For additional information contact:

Carrie Hastings

Acting Services & Access Manager
Docking State Office Building
915 SW Harrison
Suite 580-W
Topeka, KS 66612
carrie.hastings@srs.ks.gov

Karen Beckerman

Assistant Director, Benefits & Services
Docking State Office Building
915 SW Harrison
Suite 580-W
Topeka, KS 66612
karen.beckerman@srs.ks.gov

Clarence Small

Administration for Children and Families
601 E. 12th Street, Room 276
Kansas City, Missouri 64106
816-426-3981

INFORMATION INCLUDED FOR EACH INSTRUMENT

Section 3



The purpose of this resource document is to provide information, in one place, about existing screening and assessment instruments designed for use with children under age 3 and their families, as well as instruments designed for assessing services provided by programs serving them. Thus, we cast a broad net and include a wide range of screening and assessment tools of potential use to programs. Many of the instruments described are established instruments that yield a standard score that places the child's performance in the context of other children of the same age. We also include some data collection tools that may be useful, such as implementation rating scales and questionnaires that include questions on family practices, health, and health care receipt from the national Early Head Start Research and Evaluation Project (EHSREP).

We did not set strict inclusion criteria, but tried to provide information on a range of features for each instrument so programs can make informed decisions in selecting instruments. Each program must determine the

purposes for considering a particular instrument and evaluate how well the instrument fulfills those purposes.

In general, because of their limited applicability for programs serving infants and toddlers, we did not include measures for which the lowest appropriate age for administration was older than 2 years. We made an exception for certain instruments, such as the Woodcock Johnson III and Peabody Picture Vocabulary Test, that Head Start programs sometimes use and that may be helpful for continuity when children go on to Head Start.

We consulted multiple sources of information to identify instruments for inclusion in this resource document. For the 2003 report, we looked at the National Early Head Start Research and Evaluation Project (EHSREP) to identify instruments used by the national and local research teams and instruments that research programs used. We held group discussions with Early Head Start program staff at the 2002 Birth to Three Institute to learn about screening and assessment tools they are using in their programs. Information was provided about screening and assessment tools that Early Head Start programs are currently using. We consulted with researchers and technical assistance experts. Finally, we conducted a literature review to identify instruments that are used widely and have been developed and/or normed within the past 15 years, or after 1987. For the 2011 report, we updated current profiles of instruments if the

publisher had made newer editions available by April 2010. We also added instruments that were used in the Early Head Start Family and Child Experiences Survey (Baby FACES) and were more widely used among Early Head Start programs in recent years. We developed a list of 39 instruments, ranked them as high and lower priority, and gathered feedback from Baby FACES researchers and technical assistance experts to determine whether additional instruments should be considered or if currently listed instruments should be omitted. This process resulted in the selection of 29 measures (16 profiles from the 2003 report for updating; 13 new measures requiring development of a profile).

The instruments included in this document were developed for a variety of purposes and by individuals from different disciplines. Thus, you may find that some instrument names are overly technical or offensive. In these cases, you may want to present the instruments to parents using a less technical name that describes what the instrument measures in terms that parents will understand. For example, you might want to refer to the Parent-Child Conflict Tactics Scale as a questionnaire on discipline and responses to children's behavior.

The screening and assessment instruments in this resource document are presented in three groups: (1) instruments for measuring child development; (2) instruments for measuring parenting, the home environment, and parent well-being; and (3) instruments for measuring

program implementation and quality. Within each group, instruments are in alphabetical order. Summary tables listing the instruments are presented at the beginning of each group of instruments.³ This resource document is intended to be a living document that will be updated as new screening and assessment instruments are identified or become available. The first update to this resource document occurred in 2011.

We gathered information about each instrument from different sources, depending on the type of instrument. For the more formal, copyrighted instruments, we relied primarily on the manuals or Web-based information available from the authors or their publishers. If we found a key research article about a formal instrument, we also reviewed it and included the pertinent information. For the more experimental, less formal instruments, we reviewed the instrument itself and the supporting material we were able to locate, such as research reports and published articles, reviews conducted by others, and personal communications with authors of the instruments. For some instruments, information included in the 2011 update is drawn from findings from the Baby FACES study. These are instruments that are more exploratory or that have not been used widely in research or in Early Head Start programs. Each entry includes a

³ THE INCLUSION OF AN INSTRUMENT IN THIS RESOURCE DOCUMENT DOES NOT CONSTITUTE ENDORSEMENT OF THE INSTRUMENT BY THE AUTHORS, MATHEMATICA POLICY RESEARCH, OR THE U.S. GOVERNMENT.

reference section that identifies the sources of information we used.

Many of these instruments are grounded in developmental theory and research. Developers of standardized tests for children usually begin with their theory of how abilities develop and identify areas to be assessed. Then they create items to measure the identified areas and try them with children to determine whether the items discriminate among children by age. After a core set of items is identified, test developers often launch a large, nationally representative study to test the items and obtain statistical information about how the study participants performed on each item. From the study findings, the test developers determine the best set of items, develop rules about where to begin and end the test, and decide on procedures for converting raw scores (based on summing the number of items answered correctly or on the average rating across items on a rating scale) to norm-referenced scores. The norm-referenced scores take advantage of the nationally representative study and allow comparisons between how an individual child performed on the test and how children of the same age in the study performed. The nationally representative study also provides information about how the instrument works with diverse and low-income populations.

Other types of research also provide important information about a screening or assessment instrument. Studies that use a new instrument in conjunction with established

instruments that measure the same ability or skill provide information about whether the new instrument measures what it was intended to measure. Other studies compare how well the new instrument predicts children's performance in a given skill area many years later. Because they take a long time to conduct, these studies are not available for very new instruments, but they can be valuable in evaluating an instrument administered when children are young.

No screening or assessment instrument performs perfectly across all the dimensions practitioners and researchers believe are important (such as the statistical properties of the instrument or how easily the resulting information feeds back into individualized intervention planning) and for all the purposes for which the instrument may be used. We encourage you to weigh the information described for each instrument according to your program's theory of change, your comprehensive plan for gathering and analyzing data, and the purposes for which you will use the information. Consultation with an expert may help you sort through this information and select screening and assessment instruments.

The language that describes screening and assessment instruments is filled with jargon. Box 4 defines the key terms used in this document.

The rest of this chapter includes a summary of what you will find described for each instrument included in Appendix C of this resource document. Each entry includes a summary table and a more detailed description

of the topics we identified as most useful for making comparisons across instruments. Note, profiles of instruments updated and added in 2011 tend to be longer and include more descriptive information than the 2003 profiles. Each updated and new profile followed the descriptive guidelines below; however, for all but the Other Languages section we did not specify if information was not available. The topics in the summary table include:

- ***Authors, Publisher, Ordering, and Initial Material Cost Information.*** This information will allow you to obtain the instruments. Instruments updated and added in 2011 include a link to the publisher's website and the instrument itself when available. Some publishers will provide an inspection copy of the materials for a short period of time at no charge. Some publishers require that only trained psychologists or other assessment professionals purchase and use the materials, because the content of the instruments must be kept confidential and the instruments must be administered and used in accordance with professional guidelines. We list the cost for the initial materials required to use the instruments. For some copyrighted materials, you will be required to purchase a score sheet for each screening or assessment you conduct. You may be able to negotiate with the publisher for a reduced price if you are buying in bulk.

Box 4

BRIEF DEFINITIONS OF KEY TERMS

Assessment. Assessment is a generic term referring to a variety of procedures for obtaining systematic information on a child's, parent's, family's, or program's strengths or needs. As noted in Chapter I, the Head Start Program Performance Standards focus on the child and family assessment purposes of identifying "(i) the child's unique strengths and needs and the services appropriate to meet those needs; and (ii) the resources, priorities, and concerns of the family and the supports and services necessary to enhance the family's capacity to meet the developmental needs of their child." These two major purposes of assessment are sometimes described as providing information for individual diagnosis and program planning. The purposes of a diagnostic assessment are to (1) identify whether an individual has special needs, (2) determine what the problems are, (3) suggest the cause of the problems, and/or (4) propose strategies to address the problems (Meisels and Provence 1992). The purposes of an assessment for program planning are to (1) learn about an individual's ability to perform particular tasks or achieve mastery of particular skills, and (2) design intervention activities for the individual that support the completion of tasks and mastery of skills over time. Depending on the purpose of the assessment process, it may include norm-referenced tests; observations in the home, child care, early intervention, program, or school setting; interviews with family members, child care providers, or others who may provide important information about the individual; and ratings by adults knowledgeable about the child (including a parent, caregiver, or teacher) (Sattler 1992). The performance standards also require programs to conduct an "assessment of community strengths, needs, and resources," as well as an annual program self-assessment of "effectiveness and progress in meeting program goals and objectives and in implementing federal regulations."

Screening. Screening is made up of a set of activities designed to identify individuals who have a high probability of exhibiting delayed, abnormal, or problematic development. The screening is intended to identify problems at an early stage and to use this information to flag individuals for further, in-depth assessment activities.

Basal. A basal is established on a standardized test when the individual demonstrates that he or she successfully completes the first few items administered. On most standardized tests, the tester begins administering the items based on how old the individual is, starting later if the individual is older. If the individual passes the number of items specified in the test manual for establishing a basal, the tester is able to assume that the individual would have gotten all of the previous items correct and adds in the number of untested items to the correctly passed items administered to the individual. If the individual does not pass the specified number of items, the tester would administer earlier items until the prescribed number of items are passed or the tester reaches the start of the test. Using a basal rule saves time during the testing session and reduces fatigue.

Ceiling. A ceiling is established on a standardized test when the individual demonstrates that he or she fails a few of the later items administered. On most standardized tests, the tester continues administering the items until a certain number (either in a row or a proportion, such as six out of eight in a row) are failed. If the individual fails the number of items specified in the test manual for establishing a ceiling, the tester ends the test and is able to assume that all later test items would be failed by that individual as well. This saves time during the testing session and reduces fatigue.

Criterion-Referenced Test. This type of test compares an individual's performance to an established measure of performance rather than to the performance of others. Criterion-referenced tests will usually include a measure of mastery, or how well a child is able to complete a task. For example, if a test required that a child identify all of the letters of the alphabet, that would be a criterion-referenced test. We would be able to describe the child's mastery of the test by using statements such as, "The child is able to identify 80 percent of the letters in the alphabet."

Norm-Referenced Test. This type of test compares an individual's performance to the performance of others on the same measure. Usually, the norms are developed from data collected from a large, nationally representative group of individuals.

Box 4

BRIEF DEFINITIONS OF KEY TERMS (CONTINUED)

Reliability. Indicators of reliability tell how dependable an assessment or screening tool is for the purpose it is used. Reliable tools are stable over time and include items that measure the same thing in different ways. For tools that require standardized observation (for example, child care quality observations or ratings of children's behavior), the scores obtained by two different, well-trained observers must be similar to be considered reliable. Statistical measures of reliability are typically reported as correlation coefficients, which range from 0 to 1.0, with a higher value reflecting greater reliability. Many researchers and test developers require that assessment and screening tools have reliability values of 0.7 or higher. For our summary descriptors, we adopted a criterion of 0.65, which reflects a rule of thumb commonly used in the field.³ Typical indicators of reliability include measures of consistency of results and stability over time:

- **Internal consistency.** If the individual items in an instrument tool measure the same thing (for example, they all assess motor ability or language development), the measure is considered to be internally consistent. One measure of internal consistency is split-half reliability. To demonstrate this, test developers and researchers test a group of individuals, then split the test items in half, usually by grouping the odd- and even-numbered items. If the two groupings of the test items are highly correlated with each other, the split-half reliability is considered to be acceptable. Another measure of internal consistency reliability is based on the correlations among all of the individual test items. This index of internal consistency is called Cronbach's alpha (named after the researcher who developed the statistical formula)

Stability. By this measure, an assessment is reliable to the extent the procedure yields the same result on two different occasions. Test-retest reliability involves testing the same group of individuals at least twice, with a relatively short interval between assessments, usually no longer than a few days or weeks apart. The higher the test-retest reliability, the more stable the assessment tool is considered to be. Longer periods between administrations of the same assessment will reduce the reliability, partly because the individual's situation (for example, skill) can be expected to

change. Some assessment tools have two versions of the same test so that the same skills or behaviors can be assessed a second or third time (as in a pre-post or longitudinal study). In such cases, test developers include information on alternate form reliability. To demonstrate that both forms of the test are essentially equivalent, a random half of a large group of individuals is given one form of the test and the other half is given the other form. Alternate form reliability is demonstrated if the scores of the two groups are highly correlated.

Reliability of administration. Another reliability consideration applies to assessment tools that require an observer to score a child's or parent's behavior or complete a rating or checklist describing the behavior observed. To use such assessments in evaluation, researchers and test developers want to be sure that these ratings can be made consistently. One index of consistency is the extent to which two trained observers obtain the same scores when they do their observations at the same time, although independently. This index is referred to as inter-rater reliability. It is usually reported either as the correlation between the scores or ratings obtained by the two observers or as the percentage of items on which the two agree.

Representativeness of Norming Sample. Standardized screening and assessment tools provide information about how the children and parents in your program are doing compared to the group (or sample) of individuals the test developers or researchers included in their norming group. Knowing whether the norming sample was nationally representative or representative of the children or parents in your program is important in deciding whether to use a screening or assessment tool. Most test authors include this information in their manuals. In general, it is better if the norming sample includes individuals of the same age group that you will be assessing, as well as geographic and racial/ethnic diversity, so that the assessment results will be relevant to the families in your program.

Box 4

BRIEF DEFINITIONS OF KEY TERMS (CONTINUED)

Validity. Indicators of a screening or assessment tool's validity provide information about whether the tool measures what it is supposed to for the purpose it is being used. Several types of validity are commonly used:

- **Content validity.** This indicator of validity provides information about whether the screening or assessment tool includes items that are a good representation of the area the tool is supposed to measure. There are no statistics associated with content validity. Instead, it is based on professional judgment from reviews of the items to verify that what they are measuring represents the domain of development that the developer intended them to measure and that they provide variety and a range of difficulty. A good manual will include a description of the procedures followed in ensuring that the content is appropriate and representative.
- **Criterion-related validity.** Criterion-related validity indicates how well performance on the screening or assessment tool compares with a criterion, or an independent measure of what the assessment is designed to predict. The criterion measure can be obtained at about the same time or after some interval:
 - To establish **concurrent validity**, test developers and researchers administer the new screening or assessment tool as well as a similar, established tool to the same individuals within a few hours or days. If the correlation between the two measures is high, concurrent validity is established. Strict interpretations require concurrent validity to reach levels of .70 or higher, but as a rule of thumb, many researchers accept .50 or higher as acceptable.³ Sometimes concurrent validity is expressed in terms of percent agreement between the two measures. In this compendium, we consider 80 percent agreement or higher as acceptable.
 - To establish **predictive validity**, researchers and test developers determine whether the screening or assessment tool conducted at one time point with a group of individuals is correlated with later functioning (these studies are often conducted over two to five years or more). If the correlation between the two measures obtained across the time interval is high, predictive validity is established. If, for example, a measure of vocabulary at age 3 is highly correlated with a test of reading ability in second grade, the vocabulary test could be said to have predictive validity. In some cases, researchers use other activities or events as the criterion, rather than another assessment. For example, predictive validity might be established by correlating age 3 vocabulary with children's second-grade language report card grades. In general, the younger the child being assessed, the poorer the predictive validity. There is a long history of poor predictive validity among infant tests, with almost none meeting high levels of validity, such as .80. Researchers have advanced many explanations for this, including the important contributions of the different environments to which children are exposed. Because we know the predictive validity of infant and toddler assessment tools is low, in this compendium, we consider a correlation of .40 to be adequate for establishing predictive validity.

Box 4

BRIEF DEFINITIONS OF KEY TERMS (CONTINUED)

Scoring.¹ Alone, the scores from screening and assessment instruments (raw scores) have limited value. It is only when they are compared against a similar group (or norming sample) of children with known characteristics that a child's score becomes meaningful. Because of this, instrument developers often provide the user with tables for converting raw scores into scores that are normed to a comparison sample. Below are some of the more frequently used normative scores:

- **Percentile rank.** The percentile rank indicates a score's relative ranking, in units 0 to 100, to other scores in the norming sample. A child whose score is at the 65th percentile has scored higher than 65 percent of the children in the norming sample. However, percentiles are not easily comparable to each other because the raw score difference between percentiles will vary depending on the percentiles' location. The raw score differences between percentiles at the extreme ends of the percentile distribution are larger than raw score differences in the middle of the percentile distribution.
- **Stanine score.** Like percentile ranks, stanine scores provide information on children's performance relative to children in the norming sample, but without the restriction on comparing scores. Stanines divide the normal curve into nine intervals, with the lowest scores falling into the first stanine, the highest scores falling into the ninth stanine, and the fifth stanine straddling the midpoint of the distribution. Except for the two extreme stanines (the first and the ninth), each stanine is one-half of a standard deviation unit, and equal differences between two pairs of stanines represent equal differences in performance. A disadvantage of stanine scores is that they magnify small differences between raw scores that fall on either side of a point separating adjacent stanines.
- **Standardized score.** Standardized scores express the difference between a raw score and the mean score in standard deviation units.² Standard scores have the properties of the normal curve and maintain the absolute differences between the raw scores. Thus, the difference in performance between standard scores of 85 and 90 is the same as the difference between standard scores of 55 and 60. Three types of standard scores are often used: T-scores, quotients, and normal curve equivalents (NCEs). T-scores have a mean of 50 and a standard deviation of 10, while quotients have a mean of 100 and a standard deviation of 15, and NCEs have a mean of 50 and a standard deviation of 21.06. Most tests of cognitive abilities have a mean of 100 and a standard deviation of 15. For most standardized tests, we consider scores within 30 points of the mean (from 70 to 130) to be in the "normal" range.
 - **Age-equivalent scores.** An age-equivalent score is the average raw score of children at that age in the norming sample. The age-equivalent score corresponding to a child's raw score provides information on the child's level of performance in terms of the age at which that level of performance could be expected, based on the performance of children in the norming sample.
 - **Sensitivity** is a measure of an instrument's ability to correctly identify persons with the disorder as having the disorder.
 - **Specificity** is a measure of an instrument's ability to identify persons who do not have the disorder as not having the disorder.

¹ This discussion is important for interpreting scores from standardized instruments. Scores from other instruments can also be interpreted meaningfully if you can compare the performance of children or parents across two points in time (such as comparing scores at the beginning and end of their program experience).

² A standard deviation is a measure of the score's dispersion or variability in a sample. The proportion of scores within a standard deviation unit of the mean score is known. For example, in a normal distribution, 68 percent of all the scores fall between one standard deviation below and one standard deviation above the mean. Thus, scores expressed in standard deviation units enable the user to understand how a child has performed relative to other children in the sample.

³ Instruments updated and added in 2011 present uncorrected correlation coefficients. We state whether corrected coefficients are presented and if available, how corrections were made.

- **Representativeness of Norming Sample.** As described in Box 4, knowing whether the norming sample was nationally representative or representative of the children or parents in your program is important in deciding whether to use an instrument. Your screening and assessment plan will include the purpose for each screening and assessment. If you are interested in how the children in your program are performing compared with children nationally, you will want to choose an instrument with a nationally representative norming sample.
- Knowing how children from low-income families in the norming sample performed compared with all children nationally can also be important for interpreting assessment results. For example, the Early Head Start Research and Evaluation study found that children's standardized scores on the Bayley Mental Development Index decreased between 14 and 24 months of age and remained at the 24-month level at 36 months. This pattern has also been found in other studies of low-income children and in the Bayley norming sample. In this case, the decrease in standardized scores reflects differences in the composition of the test at different ages. At 14 months of age, the Bayley does not include many items directly focused on language development. At 24 and 36 months, the Bayley includes many items that tap language development. The decrease in standardized scores among low-income children as they get older indicates that low-income children score less well compared to children nationally as language development becomes a more important part of the test.
- **Languages.** We included the languages in which the instruments are available. Some instruments have unofficial translations used in the field, but we restricted our listing to the languages that are available from the authors or publishers. If you are planning to use an instrument to compare the children in your program with those in the instrument's norming sample, using an unofficial translation or directly translating the instrument into another language will result in scores that may not be comparable to the norming sample scores. According to the strictest standards, such scores are not valid.
- **Type of Instrument.** We categorized the child and parent instruments as one of three types: (1) direct child or parent instruments, in which a trained individual works one-on-one with the child or parent to administer the instrument; (2) observation, in which a trained individual observes the child or parent and either rates or scores the behaviors of interest; and (3) parent report or self-report, in which the parent reports about the child or himself or herself. These basic categories apply to most of the other areas we reviewed as well, such as quality of program services. As needed, we used

different descriptors to make our meaning as clear as possible.

- ***Age Range and Administration Interval.*** We have included the age range for which the instrument is appropriate, as well as the recommended time between administrations of the instrument, if given. Some instruments are designed to be administered at regular intervals, and that information is also included.
- ***Personnel, Training, Administration, and Scoring Requirements.*** We described whether the instrument requires administration by a consultant or expert with clinical training, a highly trained program staff member, or a clerical program staff member. We included an estimate of how much time a person at the level required would need to learn, conduct, and score the instrument. Some of the authors and publishers suggest that trainees have an administration reviewed by an experienced assessor. If so, we also included this requirement. Some of the authors and publishers offer group training on the use of their instruments, and we included that information and the cost of the training, if it is available.
- ***Summary.*** We chose five key features of the instruments to include in the summary table. Each feature has descriptors numbered from 1 to 3. A descriptor of 1 indicates a lack of information or lower-level performance on the feature, a descriptor of 3

indicates a higher-level of performance, and 2 is intermediate. We include this summary section to help you compare the features of the instruments, but do not consider this information as a recommendation of one instrument or another. Only you and your staff can decide which features are most important to you. The purposes of your screening and assessment must guide your choices about which instruments to use. The features we include in the summary section are:

- **Initial material cost:** 1 (under \$100), 2 (\$100 to \$200), 3 (more than \$200).
- **Reliability:** 1 (none described); 2 (all or mostly under .65); 3 (all or mostly .65 or higher). See Box 4 for a brief definition of the various types of reliability. We chose these groupings based on the prevalent rule of thumb researchers and assessment developers use. Other things being equal, the higher the reliability is, the better the instrument is.
- **Validity:** 1 (none described); 2 (all or mostly under .5 for concurrent; all or mostly under .4 for predictive); 3 (all or mostly .5 or higher for concurrent; all or mostly .4 or higher for predictive). See Box 4 for a brief definition of the various types of validity. We chose these groupings based on the prevalent rules of thumb researchers and instrument developers use. Generally, the higher the validity is, the better. It is especially

challenging to create instruments for infants and toddlers that strongly predict how the children will do as preschoolers. Therefore, the grouping for predictive validity reflects a less stringent criterion for the highest grouping.

- **Norming sample characteristics:**
1 (none described); 2 (older than 15 years, not nationally representative or representative of the low-income population enrolled by Head Start programs serving infants and toddlers); 3 (normed within past 15 years, nationally representative or representative of the low-income population enrolled by Head Start programs serving infants and toddlers). See Box 4 for a brief definition of representativeness of the norming sample. This section also includes information on the date that the norming sample was obtained. The more time that has elapsed since the norming sample was obtained, the less likely it is to be representative. Many authors/publishers re-norm their assessments every 10 to 12 years to keep them up-to-date. We chose 15 years as the critical time here.
- **Ease of administration and scoring:**
1 (not described); 2 (self-administered or administered and scored by someone with basic clerical skills);

3 (administered and scored by a highly trained individual). The administration and scoring requirements for each instrument vary and these descriptors help you determine what is involved for these steps.

The other topics included for each instrument are:

- **Description.** This section provides an overview of what the instrument was designed to measure, the age range of individuals it may be used with, the number of items, how it is administered, and what types of information can be derived (including any scores and subscale scores). Instruments updated and added in 2011 include information about the stimuli and manipulatives used to aid administration. They also include a description of changes from the previous version of the instrument.
- **Other Languages.** Instruments updated and added in 2011 include this section to name and describe translations of the instrument. Detail on the norming sample, reliability, validity, and English language equivalence or comparability is provided when available and noted when it is unavailable. Unofficial translations of the instrument that are used in the field are not included.
- **Uses of Information.** To help you match your intended purposes for an instrument with the results, we included a summary of how the information that comes from an instrument may be used. Some of the

instruments are clearly designed for screening children, some for in-depth assessment, some for allowing comparisons to a national norming sample, some for parent education, and some for feeding back into individual intervention planning and continuous program improvement.

- **Reliability.** Indicators of an instrument's reliability help determine whether an instrument is dependable. For example, a dependable instrument is also stable, and the results would be similar if the instrument was administered to the same individual several times in a short period. Box 4 summarizes key information about what to look for in reports of an instrument's reliability. The types of reliability summarized in the resource document entries include:
 - **Measures of internal consistency** (split-half reliability, internal consistency reliability) that indicate the extent to which the items in the instrument “hang together” and tell a coherent story about the child or adult's functioning
 - **Measures of stability** (test-retest reliability, alternate form reliability) that indicate the extent to which the instrument yields the same results when used at different times or using a different form of the instrument (for those that have multiple forms)
 - **Measures of the reliability of administration** (inter-rater reliability)

that indicate the extent to which two different observers or instrument administrators would interpret and record the information in the same way

- **Validity.** Indicators of an instrument's validity help determine whether the instrument really measures what it is supposed to for the purpose it is being used. For example, if an instrument is supposed to provide an estimate of a toddler's language production, how the child performs on the instrument should be similar to how the child performs on another established instrument of language production. We summarize key information about what to look for in reports of an assessment's validity in Box 4. The types of validity summarized in the resource document entries include:
 - **Content validity**, which relies on expert judgment to determine that an instrument actually measures what it is intended to measure. Instruments updated and added in 2011 also include information related to literature reviews and factor analysis of items.
 - **Criterion-related validity**, including concurrent validity, which indicates how well the instrument results relate to other information collected at the same time, and predictive validity, which indicates the extent to which the instrument results are related to later functioning.

- **Bias Analysis.** Instruments updated and added in 2011 include this section describing tests for whether instruments perform differently among subgroups (e.g., age, race/ethnicity, or spoken language) when differences are not expected. The results of differential item functioning (DIF) analysis are included here. DIF examines whether subgroup responses to particular items are significantly different, and instrument developers may opt to retain or remove items in question.
- **Method of Scoring.** Child screening and assessment instruments may be scored using a simple pass/fail point system, or they may use a broader range of response categories, such as whether the child usually exhibits a particular behavior, is just starting to show the behavior, or does not yet display the behavior. In this section, we summarize the response categories used in the instrument and the types of scores it is possible to compute.
- **Interpretability.** Many instrument authors and publishers provide information about how to interpret what a score or range of scores means as to whether the child is functioning at the level expected for his or her age or whether additional information may be needed. These guidelines are helpful in making sense out of the results. In this section, we summarize what is available to help you interpret the information that comes from each instrument.
- **Training Support.** In this section, we summarize what training in the use of the instrument the authors and publishers recommend. We also describe training materials, products, or sessions available. Some authors and publishers include a lot of information about how to prepare to administer their instruments, while others provide little. Some provide training videotapes or exercises as part of the purchase of the instrument. In this section, we summarize what the authors and publishers include to help you identify who needs to administer the instrument and the resources available for training them.
- **Adaptations/Special Instructions for Individuals with Disabilities.** Some instruments are designed specifically to assess the abilities or performance of individuals with disabilities, but most are not. In this section, we describe adaptations or instructions the authors or publishers included for working with people with disabilities.
- **Report Preparation Support.** Some instruments include summary sheets or software to help you prepare individual reports based on the results. These reports may be designed to help you customize the program for a given child or parent or to help you share information with parents. Some instruments also include recommendations on how to present reports to parents.

- **References.** In this section, we give the full citations for the instruments, manuals, and other sources of information we used to complete each entry. We also include citations for any other materials the authors/publishers make available about the instrument, such as training videotapes and computer scoring programs.

The entries are organized alphabetically in three groups: (1) measures of child development; (2) measures of parenting, the home environment, and family well-being; and (3) measures of program implementation and quality. In front of each group of entries is a summary table that lists the instruments profiled in that section and summarizes their main features.

References



Administration for Children and Families. *Information Memorandum: Using Child Outcomes in Program Self-Assessment*. Publication no. ACYF-IM-HS-00-18. Washington, DC: U.S. Department of Health and Human Services, August 2000.

Administration for Children and Families. *Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start. Volume I: Technical Report*. Washington, DC: U.S. Department of Health and Human Services, 2002.

Administration on Children, Youth and Families. *Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families. Volume I: Technical Report*. Washington, DC: U.S. Department of Health and Human Services, 2001a.

Administration on Children, Youth and Families. *Head Start FACES: Longitudinal Findings on Program Performance, Third Progress Report*. Washington, DC: U.S.

Department of Health and Human Services, 2001b.

Administration on Children, Youth and Families. *Leading the Way: Characteristics and Early Experiences of Selected Early Head Start Programs. Volume I: Cross-Site Perspectives.* Washington, DC: U.S. Department of Health and Human Services, 1999.

Andreassen, C., and P. Fletcher. *Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) Psychometric Report for the 2-Year Data Collection* (NCES 2007-084). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2007.

Bayley, N. *Bayley Scales of Infant Development, Second Edition, Manual.* New York: The Psychological Corporation, Harcourt Brace & Company, 1993.

Boller, K., S. Sprachman, and the Early Head Start Research Consortium. *The Child-Caregiver Observation System Instructor's Manual.* Princeton, NJ: Mathematica Policy Research, 1998.

Bowman, Barbara T., M. Suzanne Donovan, and M. Susan Burns. *Eager to Learn: Educating Our Preschoolers.* Washington, DC: National Academy Press, 2000, pp. 233-260.

Buss, A.H., and R. Plomin. *Temperament: Early Developing Personality Traits.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1984.

Centers for Disease Control and Prevention. National Health Interview Survey. Available at <http://www.cdc.gov/nchs/nhis.htm>. Accessed July 9, 2011.

Early Head Start National Resource Center. *Developmental Screening, Assessment, and Evaluation: Key Elements for Individualizing Curricula in Early Head Start Programs.* Technical Assistance Paper #4. Washington, DC: ZERO TO THREE, undated. Available at [http://eclkc.ohs.acf.hhs.gov/hslc/hs/resources/ECLKC_Bookstore/PDFs/FinalTAP\[1\].pdf](http://eclkc.ohs.acf.hhs.gov/hslc/hs/resources/ECLKC_Bookstore/PDFs/FinalTAP[1].pdf).

Gilliam, Walter S., and Linda C. Mayes. "Clinical Assessment of Infants and Toddlers." In *Child and Adolescent Psychiatry: A Comprehensive Textbook (3rd Edition)*, edited by M. Lewis. Baltimore, MD: Lippincott, Williams, and Wilkins, 2002.

Kontos, S., C. Howes, M. Shinn, and E. Galinsky. *Quality in Family Child Care and Relative Care.* New York: Teachers College Press, 1995.

Luster, T., K. Rhoades, and B. Haas. "The Relation between Parental Values and Parenting Behavior: A Test of the Kohn Hypothesis." *Marriage and the Family*, vol. 51, 1989, pp. 139-147.

Martin, Sue. *Take a Look: Observation and Portfolio Assessment in Early Childhood (2nd Edition).* Reading, MA: Addison-Wesley Publishing Company, 1999.

McHale, Jean, Colleen O'Neal, So-Yun Lee, and Lisa Berlin. High Chair (Frustration) Assessment. Unpublished document, 1999.

Meisels, Samuel J., and Sally Provence. *Screening and Assessment: Guidelines for Identifying Young Disabled and Developmentally Vulnerable Children and Their Families*. Washington, DC: ZERO TO THREE, 1992.

Pearlin, L.I., and C. Schooler. "The Structure of Coping." *Journal of Health and Social Behavior*, vol. 22, 1978, pp. 337-356.

Pianta, R.C. *Student-Teacher Relationship Scale [STRS]*. Lutz, FL: Psychological Assessment Resources, Inc., 2001.

Ross, Catherine E., James Mirowsky, and Joan Huber. "Dividing Work, Sharing Work, and in Between: Marriage Patterns and Depression." *American Sociological Review*, vol. 48, no. 6, 1983, pp. 809-823.

Shaw, Daniel S., Kate Keenan, and Joan I. Vondra. "The Developmental Precursors of Antisocial Behavior: Ages 1-3." *Developmental Psychology*, vol. 30, 1994, pp. 355-364.

Shepard, Lorrie, Sharon Lynn Kagan, and Emily Wurtz (eds.). *Principles and Recommendations for Early Childhood Assessments*. Submitted to the National Education Goals Panel by the Goal 1 Early Childhood Assessments Resource Group, January 1998.

U.S. Department of Health and Human Services. "Head Start Program: Final Rule." *Federal Register*, vol. 61, no. 215, November 5, 1996, pp. 57 186-57 2278.

U.S. Department of Health and Human Services. *The Statement of the Advisory Committee on Services for Families with Infants and Toddlers*. Washington, DC: U.S. Department of Health and Human Services, September 1994.

Vogel, Cheri A., Kimberly Boller, Yange Xue, Randall Blair, Nikki Aikens, Andrew Burwick, Yevgeny Shrago, Barbara Lepidus Carlton, Lara Kalb, Linda Mendenko, Judy Cannon, Sean Harrington, and Jillian Stein. *Administration for Children and Families. Learning as We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children*. OPRE Report #2011-7. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. 2011.

Wolverton, E. Dollie. "An Introduction to Developmental Screening in the Education Component." *National Head Start Bulletin*, no. 43, January/February 1993, pp. 3-4.

Zill, Nicholas, and James Peterson. *Behavior Problems Index (BPI)*. Washington, DC: Child Trends, Inc., 1985.

WORKSHEET TO HELP DEVELOP A COMPREHENSIVE PLAN FOR GATHERING AND ANALYZING DATA

Appendix A



SUMMARY OF INFORMATION COLLECTED AND SCREENING/ASSESSMENT TOOLS USED BY YOUR HEAD START PROGRAM SERVING INFANTS AND TODDLERS

Head Start programs serving pregnant women and families with infants and toddlers collect information for various purposes, using a variety of assessment tools or instruments. The table below provides a useful format for summarizing your program's current or planned use of assessment tools and instruments and their purposes.

Instrument/Tool	When Administered:						How Used:				
	At Enrollment	Scheduled Times After Enrollment	At Specific Ages	Under Certain Circumstances	Varies	Other	To Plan Individual Services	To Create Aggregate Reports For:			
								Program Planning	Funders	PIR	Other

Screening/Assessing Children's Development

Assessing Parenting or Parent Mental Health Needs

Instrument/Tool	When Administered:						How Used:				
	At Enrollment	Scheduled Times After Enrollment	At Specific Ages	Under Certain Circumstances	Varies	Other	To Plan Individual Services	To Create Aggregate Reports For:			
								Program Planning	Funders	PIR	Other
Assessing Family Social Service Needs											
Assessing Quality of Program Services											
Tracking Service Delivery											

Instrument/Tool	When Administered:						How Used:				
	At Enrollment	Scheduled Times After Enrollment	At Specific Ages	Under Certain Circumstances	Varies	Other	To Plan Individual Services	To Create Aggregate Reports For:			
								Program Planning	Funders	PIR	Other
Assessing Family Engagement/Satisfaction											
Assessing Overall Program Performance and/or Adherence to the Performance Standards											
Any Other Data Collection/Assessments											

PERFORMANCE MEASURES FRAMEWORK FOR HEAD START PROGRAMS SERVING INFANTS AND TODDLERS

Appendix B



As a national laboratory for early childhood education, Head Start has long emphasized continuous program improvement and has been a leader in developing outcomes-oriented accountability. Head Start began developing program performance measures in 1995, and in 1997 the Family and Child Experiences Survey (FACES) was launched to collect data on the performance indicators, in part to be responsive to the Government Performance and Results Act (GPRA) requirements. Following the reauthorization of Head Start in 1998, Head Start programs were required to include child outcomes in their self-assessment process by 2003. The child outcomes framework was updated in 2010 for 3- to 5-year-old children, which may be a useful resource to programs serving children from 0 to 5 years old.

In 2007, the Early Head Start Family and Child Experiences Survey (Baby FACES) was launched to collect data on a representative sample of programs and two cohorts of children (newborns and 1-year-olds). Family and child

outcomes are gathered throughout children's enrollment in the program, through age 3 (Vogel et al. 2011). Although Early Head Start programs have not been required to report child outcomes, many have started to define and measure outcomes, for several reasons. Some programs operate within Head Start programs and are doing so in conjunction with outcomes assessment in Head Start. Some programs are responding to the requirements of other funders. Yet others are acting out of a desire to improve their services to families with infants and toddlers. In Spring 2001, the Early Head Start Technical Work Group recommended that the Office of Head Start move forward to develop performance measures for Head Start programs serving infants and toddlers to support programs' efforts. The framework presented in this document resulted from an iterative process during which a wide range of stakeholders were consulted.

The Head Start performance measurement framework needed to be modified for infants and toddlers in order to be appropriate and useful for home-based, center-based, combination, and locally-designed programs, and it needed to take into account the full range of development over a much longer period, from pregnancy to age 3.

The recent completion of the national Early Head Start Research and Evaluation Project (EHSREP) also makes the time ripe for developing a new performance measures framework. The pattern of program impacts by

age 3 and variations in impacts by program approach (center-based, home-based, and mixed approaches) and key aspects of program implementation provide insights into ways that the Head Start framework must be adapted for programs serving families with infants and toddlers.

Building the Pyramid

The conceptual model underlying performance measures for Head Start programs serving infants and toddlers is based on the Head Start model (ACYF 2001), which was modified to reflect services for infants and toddlers.

The Statement of the Advisory Committee on Services for Families with Infants and Toddlers recommended four cornerstones for Early Head Start: child development, family development, staff development, and community building (Administration on Children, Youth and Families 1994). Thus, the adapted pyramid rests on these cornerstones, and the elements of management systems related to each cornerstone.

The overall goal of the base level of the pyramid is "ensuring well-managed programs that involve parents in decision-making." This goal links the key elements in management systems that correspond to the four cornerstones.

The services level of the pyramid has three main objectives: providing children with

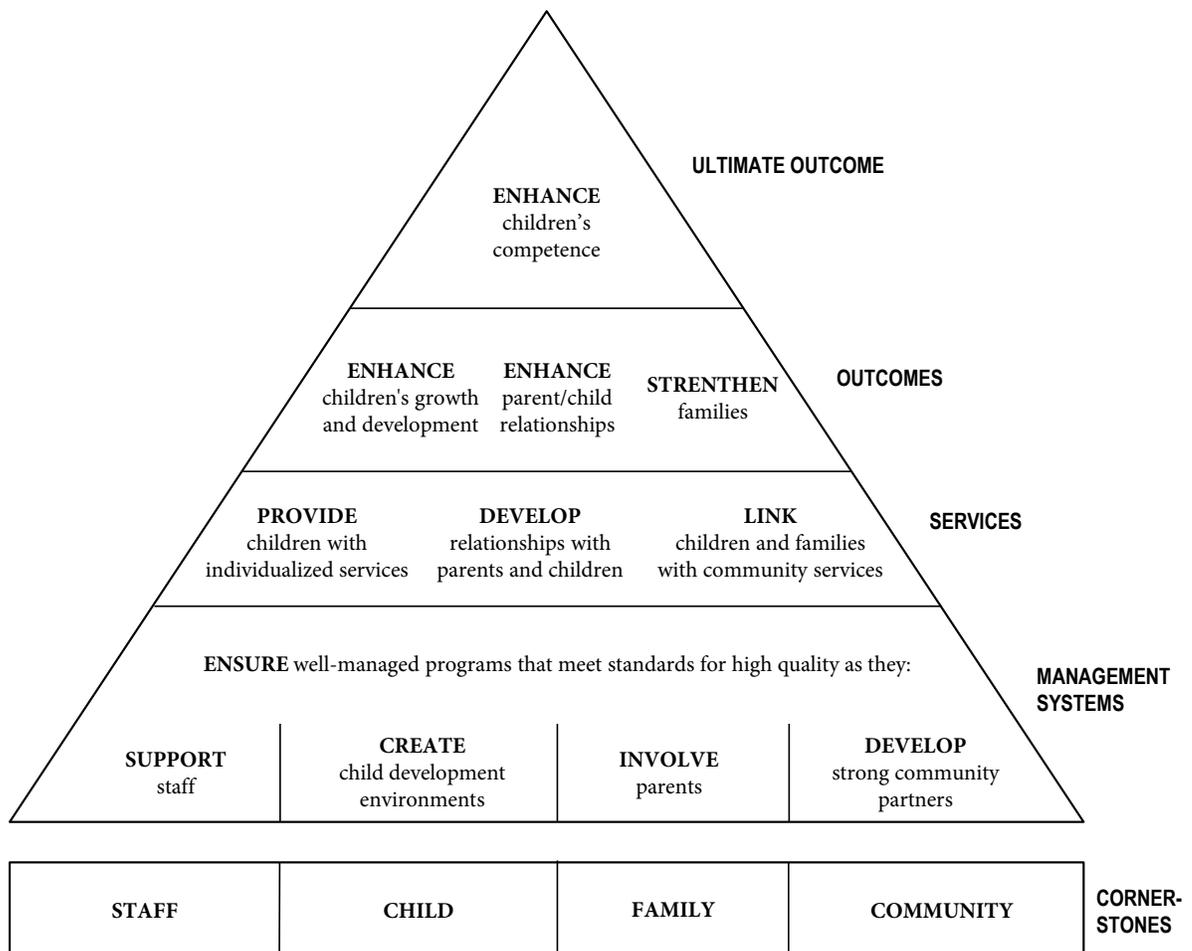
services, linking children and families to needed community services, and developing relationships with parents and children.

Similarly, on the level of child and family outcomes, there are three main objectives: enhancing children’s growth and development, strengthening families as primary nurturers of their children, and enhancing parent-child relationships. Because it is through these relationships that strengthening families can support enhanced child outcomes and the

ultimate goal of enhanced children’s competence, the objective of enhancing parent-child relationships was placed between the other two objectives on the outcomes level.

Head Start programs serving pregnant women and families with infants and toddlers may provide center-based services, home-based services, or some combination of these, reflecting different family needs, different theories of change, and different emphases on certain services and outcomes.

Figure 1 CONCEPTUAL FRAMEWORK FOR PROGRAMS SERVING INFANTS & TODDLERS



The pyramid for these programs was designed to encompass these key variations. When providing center-based services, programs emphasize the left side of the pyramid—they aim to improve children’s competence mainly by providing direct center-based services to children to enhance their growth and development (and also provide services to strengthen families and improve parent-child relationships). When providing home-based services, programs emphasize the right side of the pyramid—they aim to improve children’s competence mainly by providing services to strengthen families and improve parent-child relationships (and also work with children directly during home visits and in parent-child group socialization activities). These variations in program emphases were reflected in the differences in impacts found in the national Early Head Start Research and Evaluation Project (EHSREP) (ACF 2002).

Performance Measures Underlying the Main Objectives

Underlying the main objective in each block in the pyramid are performance measures that represent key program goals under that objective (Figure 2). The program performance measures for Head Start programs serving infants and toddlers include selected performance measures from the Head Start framework, as well as new performance measures that reflect the special features of programs serving infants and toddlers, variations in program approaches, and

lessons from the national evaluation and training and technical assistance efforts. The performance measures also reflect key requirements in the Head Start Program Performance Standards as well as elements from the Advisory Committee on Services for Families with Infants and Toddlers.

Using The Program Performance Measures Framework

The program performance measures framework for Head Start programs serving pregnant women and families with infants and toddlers is designed to support continuous program improvement efforts of individual programs, as well as those at the regional and national levels. At the individual program level, the framework, along with the program’s specific theory of change, can guide the development of plans for data collection to provide the program staff with important information on program strengths and weaknesses. This information can help focus program improvement efforts on areas where improvement may be needed most and in ways that may be most beneficial to the program. Information on strengths can be used to highlight program accomplishments and build support for the program among funders and key stakeholders. *A Resources for Measuring Services and Outcomes* is designed to support programs in these activities by providing information on creating an integrated plan and selecting measures.

References

Administration for Children and Families. Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start. Washington, DC: U.S. Department of Health and Human Services, June 2002.

Administration for Children and Families. *The Head Start Child Development and Early Learning Framework: Promoting Positive Outcomes in Early Childhood Programs Serving Children 3-5 Years Old*. Washington, DC: U.S. Department of Health and Human Services, December 2010.

Administration on Children, Youth and Families. First Progress Report on the Head Start Program Performance Measures. Washington, DC: ACYF, May 1997.

Administration on Children, Youth and Families. Head Start FACES: Longitudinal Findings on Program Performance, Third Progress Report. Washington, DC: U.S. Department of Health and Human Services, January 2001.

Administration on Children, Youth and Families. The Statement of the Advisory Committee on Services for Families with Infants and Toddlers. Washington, DC: U.S. Department of Health and Human Services, 1994.

Head Start Program Performance Measures: Second Progress Report. Washington, DC: U.S. Department of Health and Human Services, June 1998.

Vogel, Cheri A., Kimberly Boller, Yange Xue, Randall Blair, Nikki Aikens, Andrew Burwick, Yevgeny Shrago, Barbara Lepidus Carlton, Lara Kalb, Linda Mendenko, Judy Cannon, Sean Harrington, and Jillian Stein. Administration for Children and Families. *Learning As We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children*. OPRE Report #2011-7, Washington, DC. Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. 2011.

Figure 2

PROGRAM PERFORMANCE MEASURES FOR HEAD START PROGRAMS SERVING INFANTS AND TODDLERS

FOUNDATION: PROGRAM MANAGEMENT SYSTEMS

Management systems (including program governance, planning, communication, record-keeping and reporting, ongoing monitoring, self-assessment, human resources, and fiscal management) ensure well-managed programs that meet standards for high quality as they create nurturing child development environments, enhance child outcomes and promote positive parent-child interactions.

1. Programs comply with Head Start regulations.
2. Programs are well-managed operationally and financially
3. Programs design and implement services to be responsive to the needs of families in the community.
4. Programs conduct self-assessments that are used for continuous program improvement.

Management Systems ensure well-managed programs with integrated systems to support staff in working effectively with parents and children.

1. Programs employ qualified staff with the skills necessary to provide high-quality services.
2. Programs support ongoing staff development, training, and mentoring.
3. Programs support staff activities through ongoing reflective supervision.
4. Programs promote staff retention and continuity.

Ensure well-managed programs that meet standards for high quality as they develop collaborative relationships with community partners.

1. Programs form partnerships with other community programs and organizations to support an integrated community-wide response to the needs of families with young children.
2. Programs form partnerships and coordinate services with local Part C agencies.
3. Programs form partnerships and coordinate services with community child care providers to meet the needs of families and enhance the quality of local child care services through the sharing of resources, training, and knowledge.
4. Programs form partnerships and coordinate services with local health agencies and health care providers to meet the health-related needs of families.

Management Systems ensure well-managed programs that meet standards for high quality as they involve parents.

1. Parents are involved actively in program planning and decision making.
2. Programs encourage and support fathers' involvement in program planning, decision-making, and activities.

PROGRAM SERVICES

Programs develop individualized family partnership agreements with families and link children and families to comprehensive

community services in order to meet their personal goals and self-sufficiency needs.

1. Programs work collaboratively with families to identify their goals, strengths, and needed services, and offer them opportunities to develop and implement individualized family partnership agreements that take into account other family plans.
2. Programs link parents with social service agencies to obtain needed services.
3. Programs link parents with educational and employment agencies to obtain needed services.
4. Programs link parents with physical and mental health care prevention and treatment services to obtain needed care.
5. Programs link parents with needed prenatal care and education services.
6. Programs help parents secure high-quality child care in order to work, attend school, or gain employment training.
7. Programs help parents and children make a smooth transition to Head Start or other pre-school program.

Program staff develop responsive and caring relationships with parents and children

1. Staff form respectful and supportive relationships with parents through all aspects of service delivery.
2. Staff form nurturing relationships with children in group-care settings or during home visits.

3. Programs support and honor the home cultures and languages of families.

Programs provide children with age-appropriate curricular experiences and individualized educational, health, and nutritional services

1. Programs provide developmentally enriching educational environments in group-care settings and developmentally enriching parenting and child development services during home visits and group socializations.
2. Programs link children with needed medical, dental, and mental health services.
3. Programs link pregnant women with comprehensive prenatal health care and education.
4. Programs provide children in group-care settings meals and snacks that meet their daily nutritional needs, and parents receiving home-based services receive information about meeting their children's nutritional needs
5. Programs provide individualized services for parents and children, including children with disabilities.

EXPECTED OUTCOMES FOR CHILDREN AND FAMILIES

Programs strengthen families as the primary nurturers of their children

1. Parents demonstrate increased knowledge of child development and awareness of their children's developmental progress
2. Parents enhance their self-concept and emotional well-being and experience less parenting stress.
3. Parents make progress toward their educational, literacy, and employment goals.
4. Adult family members strengthen their relationships and work together in caring for children.

Programs support and enhance parent-child relationships

1. Parents demonstrate more sensitivity and responsiveness in interactions with their children.
2. Parents spend more time with their children in activities that stimulate their children's development, such as reading to their children

3. Parents provide home environments and experiences that are more supportive of their children's development.

Programs enhance children's growth and development

1. Children demonstrate improved communication, language, and emergent literacy skills.
2. Children demonstrate improved general cognitive skills.
3. Children demonstrate improved positive approaches toward learning, including improved attention skills.
4. Children demonstrate improved social behavior, emotion regulation, and emotional well-being.
5. Children demonstrate improved physical health and development.

**ULTIMATE GOAL: CHILDREN'S
COMPETENCE**

List of Measures used in the National Early Head Start Research and Evaluation Project (EHSREP) and Early Head Start Family and Child Experiences Survey (Baby FACES)

Appendix C



List of Measures

Measure	Type	Construct	EHSREP	Baby FACES
Ages and Stages Questionnaire-Third Edition (ASQ-3; Squires, Twombly, Bricker, and Potter, 2009)	CH	Screening tool for developmental delays in communication, gross motor, fine motor, personal-social, and problem solving		X
Achenbach System of Empirically-Based Assessment, Child Behavior Checklist, Aggressive subscale (Achenbach and Rescorla 2000)	CH	Aggressive behavior problems	X	
Arnett Scale of Caregiver Behavior (Arnett, 1989)	CL	Caregiver's sensitivity, harshness, and detachment	X	
Bayley Behavioral Rating Scale (Bayley 1993)	CH	Engagement of examiner, emotional regulation during assessment	X	X
Bayley Scales of Infant Development-II, Mental Development Index (Bayley 1993)**†	CH	Cognitive development	X	
Beliefs Regarding Talking and Reading scale (Luster, Rhoades and Haas, 1989)	F	Parent's beliefs about talking and reading to children	X	
Behavior Problems Index (BPI; Zill and Peterson, 1995)	CH	Parent report of child behavior		X
Brief Infant Toddler Social Emotional Assessment (BITSEA; Briggs-Gowan and Carter, 2006)	CH	Social-emotional development, behavior problems		X
Center for Epidemiological Studies - Depression Scale (CES-D; Radloff 1977)	F	Parent's depressive symptoms	X	
Center for Epidemiological Studies - Depression Scale, short form (CES-D; Ross, Mirowsky et al., 1983)	F	Parent's depressive symptoms		X (also administered to children's teachers and home visitors)
Child-Caregiver Observation System (C-COS; Boller, Sprachman, and the Early Head Start Research Consortium 1998)	CL	Quality, frequency, and types of provider-child interactions	X	
Classroom Assessment Scoring System-Toddler (CLASS-T; Pianta et al., 2010)	CL	Teacher-child interaction quality in toddler childcare classrooms		X
Composite International Diagnostic Inventories (CIDI) - Short Form, Major Depression, Generalized Anxiety Disorder, Alcohol Dependence, Drug Dependence (Nelson, Kessler, and Mroczek 1998 and World Health Organization 2002)	F	Parent's major depression, generalized anxiety disorder, alcohol abuse, and substance abuse	X	
Confusion, Hubbub, and Order Scale (CHAOS; Mathney et al., 1995)	F	Level of confusion and disorganization in child's home environment		X
Discipline Vignettes	F	Parental use of reasoning and developmentally appropriate approaches in parent-child conflict situations	X	

Measure	Type	Construct	EHSREP	Baby FACES
EAS Temperament Survey for Children (Buss and Plomin, 1984), selected items	CH	Temperament (emotionality and sociability)	X	
Early Childhood Environmental Rating Scale-Revised (ECERS-R; Harms, Clifford, and Cryer 1998)	CL	Quality of the caregiving environment in center-based care	X	
Early Communication Indicator (ECI; Luze et al., 2001; Carta et al., 2002; 2010)	CH	Expressive communication		X
Family Day Care Rating Scale (FDCRS; Harms and Clifford 1989)**†	CL	Quality of the caregiving environment in home-based care	X	
Family Environment Scale, Conflict Subscale (Moos and Moos, 2002)	F	Family conflict	X	X
High Chair (Frustration) Assessment (Shaw et al. 1994; McHale et al. 1999)	F	Parent's Sensitivity, Detachment, Negative Regard/Hostility, Positive Regard, Distance/Approach	X	
High Chair (Frustration) Assessment (Shaw et al. 1994; McHale et al. 1999)	CH	Child Distress	X	
Home Observation for Measurement of the Environment (HOME), Infant/Toddler Form (Caldwell and Bradley, 1984)**†	F	Emotional Responsivity, Maternal Verbal-Social Skills, Support of Cognitive, Language, and Literacy Environment, Absence of Punitive Interactions, Internal Physical Environment, External Environment		X
Home Observation for Measurement of the Environment (HOME), Infant/Toddler Form (Caldwell and Bradley, 2003)	F	Quality of home environment, Reading and language activities, Emotional Responsivity, Support of Cognitive, Language and Literacy Environment, Maternal Verbal-Social Skills, Absence of Punitive Interactions	X	
Home Observation for Measurement of the Environment (HOME), Preschool Form, NLSY version, selected items (Caldwell and Bradley, 2003 and Center for Human Resources Research, 2000)	F	Quality of home environment, Reading and language activities, Warmth, Support of Language and Learning, Internal Physical Environment, Harshness	X	
Home Visit Rating Scale-Adapted (HOVRS-A; Roggman et al., 2009)	S, F	Home-visiting quality and content, Visitor Strategies Quality, Visitor Effectiveness Quality		X
Infant/Toddler Environment Rating Scale (ITERS, Harms, Cryer, and Clifford 1990)**†	CL	Quality of the caregiving environment in center-based care	X	
Infant/Toddler Environment Rating Scale, Revised (ITERS-R, Harms et al., 2003)**†	CL	Quality of the caregiving environment in center-based care		X
Infant Toddler Social Emotional Assessment, Exposure to Violence Scale (ITSEA; Carter and Briggs-Gowan, 2000)**†	CH	Quality of the caregiving environment in center-based care		X
Knowledge of Infant Development Inventory (KIDI, MacPhee, 1981), selected items	F	Parent's knowledge of childrearing practices, developmental processes, and infant developmental norms and milestones	X	

Measure	Type	Construct	EHSREP	Baby FACES
MacArthur Communicative Development Inventories (CDI), Vocabulary Short Form, Level I; Part II, Section A First Communicative Gestures; and Part II, Section B Games and Routines (MacArthur CDI Advisory Board 1997; and Fenson et al. 2000a and b)**†	CH	Receptive vocabulary, productive vocabulary, use of gestures to communicate, parent-child communicative activities	X	
MacArthur Communicative Development Inventories (CDI), Vocabulary Short Form, Level II; Part II, Question on Combining Words; and Section E. Complexity (MacArthur CDI Advisory Board 1997; Fenson et al. 2000a and b)**†	CH	Productive vocabulary, complexity of speech	X	
MacArthur-Bates Communicative Development Inventory, English Toddler Short Form-Form A, (CDI; Fenson et al., 2000), Spanish Toddler Short Form (Jackson-Maldonado et al., 2003)**†	CH	Vocabulary Comprehension, Vocabulary Production		X
Nursing Child Assessment Satellite Training (NCAST) Parent-Child Interaction Program, Teaching Scales (Barnard 1994; Sumner and Spietz 1994)	F	Parent total score	X	
Nursing Child Assessment Satellite Training (NCAST) Parent-Child Interaction Program, Teaching Scales (Barnard 1994; Sumner and Spietz 1994)	CH	Child total score	X	
Child's Health Status (CDC National Health Interview Survey)	CH	Child's health status	X	
Parent-Caregiver Relationship Scale (selected items) (Elicker, et al. 1997)	F	Parent's relationship with current nonrelative provider	X	X
Parent-Caregiver Relationship Scale (selected items) (Elicker, et al., 1997)	CL	Caregiver's relationship with parents	X	
Parent-Child Interaction Rating Scales for the Two Bags Assessment-Child Scales (PCI; and Fletcher, 2007)	F	Parent behaviors, child behaviors, and dyadic interaction		X
Parental Modernity Scale (selected items) (Schaefer and Edgerton 1985)	F	Parent attitudes toward children and childrearing practices (traditional attitudes and practices; progressive attitudes and practices)	X	X
Parental Modernity Scale (selected items) (Schaefer and Edgerton 1985)	CL	Caregiver attitudes toward children and childrearing practices (traditional attitudes and practices; progressive attitudes and practices)	X	
Parenting Alliances Measure (PAM; Abidin and Konold, 1999)	F	Parents' perspective of how cooperative, communicative, and mutually respectful they are in their parenting relationship		X
Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO; Cook and Roggman, 2009; Roggman et al., 2009)	F	Positive parenting behaviors		X

Measure	Type	Construct	EHSREP	Baby FACES
Parenting Stress Index, Short Form (PSI-SF; Abidin, 1995), Parental Distress and Parent-Child Dysfunctional Interaction Subscales	F	Parental distress, Parent-child dysfunctional interaction	X	X
Peabody Picture Vocabulary Test-III (PPVT-III; Dunn and Dunn 1997) or Test de Vocabulario en Imagenes Peabody (TVIP; Dunn, Padilla, Lugo, and Dunn, 1986)	CH	Receptive language	X	
Peabody Picture Vocabulary Test-IV (PPVT-IV; Dunn and Dunn 2007)	CH	Receptive language		X
Preschool Language Scale, Fourth Edition (PLS-4; Zimmerman et al., 2002)	CH	Auditory comprehension, Expressive communication		X
Program Implementation Checklist and Rating Scales Developed for the National Early Head Start Research and Evaluation Project (ACF 2003)	CL	Implementation of program performance standards	X	X
Provider Job Commitment (Kontos et al. 1995)	CL	Provider's commitment to child care as a profession	X	
Provider Motivation (Kontos et al. 1995)	CL	Provider's motivation for beginning child care work	X	
Provider Views about Training (Kontos et al. 1995)	CL	Provider's views about the usefulness of training	X	
Puzzle Challenge (Problem-Solving) Task (Brady-Smith et al. 2001)	F	Parent Supportive Presence, Quality of Assistance, Detachment, Intrusiveness	X	
Puzzle Challenge (Problem-Solving) Task (Brady-Smith et al. 2001)	CH	Child Engagement of Parent, Persistence, Frustration with Task	X	
Three-Bag Assessment (Semi-structured Play) (NICHD Study of Early Child Care 1992; Ware et al. 1998; Brady-Smith et al. 1999 and 2000)	F	Parent Supportiveness, Sensitivity, Positive Regard, Stimulation of Cognitive Development, Detachment, Intrusiveness, Negative Regard	X	
Three-Bag Assessment (Semi-structured Play) (NICHD Study of Early Child Care 1992; Ware et al. 1998; Brady-Smith et al. 1999 and 2000)	CH	Child Engagement of Parent; Sustained Attention to Objects, Negativity Toward Parent	X	
Student-Teacher Relationship Scale (Conflict and Closeness subscales) (Pianta 2001)	CH	Conflict with child; Closeness toward child (provider report)	X	
Mastery Scale (Pearlin and Schooler 1978)	F	Parent's self-efficacy	X	
Woodcock-Munoz Language Survey: Picture Vocabulary Test (Woodcock and Munoz-Sandeval 2001)**†	F	Parental language/cognition	X	

Types are * CH - Child Measure, F - Family/Parent Measure, CL - Classroom Measure, S - Staff Measure

**† A more recent version of the instrument is included in this resource document.

Child Development Instruments



Child Development Instruments

Instrument	Screening or Assessment	Domain					Age Range	Assessment Type	Initial Material Cost	Reliability	Validity	Norming Sample	Ease of Administration & Scoring
		C	L	S-E	M	O							
Achenbach System of Empirically Based Assessment	A		X	X			1.5-5 years	3	2	3	3	3	3
Ages & Stages Questionnaires, Third Edition (ASQ-3)	S	X	X	X	X		1-66 months	3	3	3	3	3	2
Ages and Stages Questionnaires: Social-Emotional	S			X			3-66 months	3	2	3	3	2	3
Assessment, Evaluation, and Programming System for Infants and Children	A	X	X	X	X		0-3 years	2,1,3	2	3	3	1	2
Batelle Development Inventory	A	X	X	X	X		0-8 years	1,2,3	1	3	3	2	2
Bayley Scales Of Infant And Toddler Development, Third Edition (Bayley-III)	A	X	X	X	X		1-42 months	1,2,3	3	3	3	3	3
Behavior Assessment System for Children, Second Edition (BASC-2)	A		X	X		X	2-25 years	2,3	3	3	3	3	2
Brigance Inventory Of Early Development II Standardized (IED-II Standardized)	A	X	X	X	X	X	0-7 years	1,2,3	3	3	3	3	2
Carey Temperament Scales	A			X			0-12 years	3	1	2	1	2	2
Carolina Curriculum for Infants and Toddlers with Special Needs Assessment Log II	A/S	X	X	X	X		0-2 years	2	1	1	1	1	3
Denver II Development Screening Test	S		X	X	X		0-6 years	1,3	1	3	1	2	3
Developmental Observation Checklist System	S	X	X	X	X	X	0-6 years	3	2	3	2	2	3
Developmental Profile 3 (DP-3)	A/S	X	X	X	X	X	0-12 years	3	3	3	3	3	3
Devereux Early Childhood Assessment For Infants And Toddlers (DECA-I/T)	A/S			X			1-36 months	2,3	2	3	3	3	3
Early Communication Indicator (ECI)	A		X				0-3 years	2	1	3	3	3	3
Early Coping Inventory	A			X			4-36 months	2	1	3	1	1	3
Early Head Start Evaluation—Parent Interviews and Child Assessments	A/S	X	X	X	X	X	14,24,36 months	1,2,3	1	3	1	1	2,3
Early Learning Accomplishment Profile—Revised Edition	A	X	X	X	X	X	0-36 months	2	3	3	3	3	3
Early Literacy Skills Assessment (ELSA)	A	X	X				3-5 years	1	2	3	2	1	2
Expressive One-Word Picture Vocabulary Test-4 (EOWPVT-4)	A		X				2-80 years	1	2	3	1	3	3
Eyberg Child Behavior Inventory/Sutter-Eyberg Student Behavior Inventory-Revised	S			X			2-16 years	3	na	3	2	2	3
Functional Emotional Assessment Scale	A			X			7-48 months	2,1	1	3	2	1	3
Hawaii Early Learning Profile	A	X	X	X	X	X	0-36 months	1	1	1	1	1	3
High/Scope Child Observation Record	A	X	X	X			30-72 months	2	2	3	2	1	3
High/Scope Child Observation Record for Infants and Toddlers	A	X	X	X	X		1-36 months	2	2	3	3	1	3
Humanics National Infant-Toddler Assessment	A	X	X	X	X		0-3 years	2	1	1	1	1	2
Infant-Toddler Developmental Assessment	A	X	X	X	X	X	0-42 months	2,3	3	3	3	2	3

Instrument	Screening or Assessment	Domain					Age Range	Assessment Type	Initial Material Cost	Reliability	Validity	Norming Sample	Ease of Administration & Scoring
		C	L	S-E	M	O							
Infant-Toddler Social And Emotional Assessment (ITSEA) and Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	A/S			X			12-35 months	3	3	3	2	3	3
Infant Toddler Symptom Checklist	S			X	X		7-30 months	3	1	1	2	1	2
Leiter International Performance Scale-Revised	S/A	X		X			2-20 years	1	3	3	3	3	3
Macarthur-Bates Communicative Development Inventories (CDIs)-Second Edition	A/S		X				8-37 months	3	2	3	3	2	3
Mullen Scales of Early Learning	A	X	X		X		0-68 months	1	3	3	3	3	3
The Ounce Scale	A	X	X	X	X		0-42 months	2,3	2	2	2	1	2
Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4)	A		X				2.5-90 years	1	3	3	1	3	3
Preschool Language Scale, Fourth Edition (PLS-4)	A		X				0-6 years 11 months	1	3	3	3	3	3
Receptive-Expressive Emergent Language Test—2nd Ed.	S		X				0-3 years	1	1	3	1	2	3
Rossetti Infant-Toddler Language Scale	A		X				0-3 years	1,2,3	1	1	1	1	3
Temperament and Atypical Behavior Scale	S/A			X			11-71 months	3	1	3	1	3	2
Test de Vocabulario en Imagenes Peabody	A		X				2-18 years	1	2	3	1	2	3
Test of Preschool Early Literacy (TOPEL)	A	X	X				3-5 years	1	3	3	3	3	3
Toddler Attachment Sort-45 (TAS-45)	A			X		X	12-36 months	2	3	2,3	2	1	3
Vineland Adaptive Behavior Scales, Second Edition (VINELAND-II)	A		X	X	X	X	0-90 years	3	2	3	3	3	3
Vineland Social-Emotional Early Childhood Scales	A			X			0-5 years	3	1	3	3	2	3
Woodcock-Johnson III Normative Update (WJ III NU)	A	X	X			X	2 years-adult	1	3	3	3	3	3

KEY

Domains

C = Cognitive, problem-solving
L = Language, communication
S-E = Social, emotional
M = Motor, physical
O = Other

Assessment Type

1 = Direct assessment
2 = Observation
3 = Parent/Teacher self report

Initial Material Cost

1 = Under \$100
2 = \$100 to \$200
3 = More than \$200

Reliability

1 = None described
2 = Under .65
3 = .65 or higher

Validity

1 = None described
2 = Under .5 for concurrent; under .4 for predictive
3 = .5 or higher for concurrent; .4 or higher for predictive

Norming sample

1 = None described
2 = Older than 15 years, not nationally representative or representative of EHS population
3 = Normed within past 15 years, nationally representative or representative of EHS population

Ease of administration and scoring

1 = Not described
2 = Self-administered or administered and scored by someone with basic clerical skills
3 = Administered and scored by a highly trained individual

The information included in this table was drawn from the manuals or other resources available from the authors and publishers of the instruments. Individual users may have different experiences.

ACHENBACH SYSTEM OF EMPIRICALLY BASED ASSESSMENT (ASEBA), 1999

Authors:

Thomas M. Achenbach and Leslie A. Rescorla

Publisher:

ASEBA

(802) 656-8313 or 656-3456

www.ASEBA.org

Initial Material Cost:

Preschool hand-scoring starter kit: \$174

Representativeness of Norming Sample: The 1999 child behavioral checklist norming sample of 700 children is nationally representative, but restricted to children with no major physical or mental disabilities and English-speaking parents. The 1,192 caregiver-teacher norming sample is not nationally representative—989 caregivers from the 1997 norming sample augmented the 203 preschool caregivers-teachers drawn from the 1999 sample. The language development survey sample consisted of 278 parents from the 1999 sample.

Languages:

English and Spanish (child behavioral checklist only)

Type of Assessment:

Parent and caregiver report

Age Range and Administration Interval: 1.5 to 5 years. No prescribed interval, but routine use is recommended.

Personnel, Training, Administration, and Scoring Requirements:

Respondents should be able to read at the 5th grade level or higher and complete the forms in about 10 to 15 minutes. The authors recommend that a person with graduate training and familiarity with the manual interpret and score the assessment.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher, with some lower)

Validity: 3 (.5 or higher for concurrent, with some lower)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (self-administered but scored by a highly trained individual)

Description: The ASEBA uses information collected from parents and caregivers/teachers to assess the behavioral, emotional, and social functioning (including language development) of young children between the ages of 1.5 and 5 years. To get a better understanding of how the child functions under different conditions, it is recommended that information be collected from more than one adult. The ASEBA consists of two self-administered reporting forms. The parent report consists of a 99-item child behavioral checklist (CBC) and a language development

survey (LDS) that asks parents to provide the child's best multi-word phrases and words the child uses from a list of 310 words. The caregiver/teacher report (CTR) consists of a 99-item checklist similar to the CBC except 17 family-specific items have been replaced with group situation items. The 99 items in the CBC are organized into seven syndromes and two broader groupings of syndromes, while the 99 items in the CTR are organized into six of the CBC syndromes and the two broader groupings:

- Internalizing
 - Emotionally reactive
 - Anxious/depressive
 - Somatic complaints
 - Withdrawn
- Externalizing
 - Attention problems
 - Aggressive behavior
- Ungrouped (CBC only)
 - Sleep problem

The items are also organized into five DSM (Diagnostic and Statistical Manual of Mental Disorder)-oriented scales.

Uses of Information: The results can be used to structure interviews with parents, identify areas for intervention, and evaluate intervention outcomes.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): the alphas for the CBC scales ranged from .66 to .92 for the syndromes and .63 to .86 for the DSM-oriented scales. The alphas were .89 and .92 for the two broader groupings (internalizing and externalizing syndromes) and .95 for the total score. The alphas for the CTR syndromes ranged from .52 to .96 and for the DSM-oriented scales from .68 to .93. The alphas were .89 and .96 for the internalizing and externalizing groupings and for the total score, .97. (2) Test-retest reliability, with an eight-day interval between tests: the correlations were .85 and .76 for the CBC and CTR, respectively. Test-

retest studies on the LDS reported correlations greater or equal to .90.

Validity: (1) Concurrent validity: The CBC correctly classified 84 percent of a sample of children (some of whom were diagnosed as having emotional/behavioral problems), and the CTR correctly classified 74 percent of the children. Studies reported correlation coefficients between the CBC problem syndromes and the Toddler Behavior Screening Inventory and the Infant-Toddler Social and Emotional Assessment ranging from .48 to .70. In 11 studies that compared parent LDS scores with those obtained by trained examiners using other measures, the correlations between the parent's score and the trained examiner's ranged from .56 to .87. Other studies found the level of LDS agreement with other measures of language development ranged from .47 to .94. (2) Predictive validity: An 11-year longitudinal study found that children identified by the LDS to have language development problems were more likely to have weak verbal skills at age 13.

Method of Scoring: ASEBA can be hand or computer scored. Respondents complete the CBC and CTR by circling one of three responses and the LDS by circling the words the child uses spontaneously. The behavioral raw scores are derived by summing the response item values (0, 1, or 2) for the syndrome scale, syndrome groupings, and total score. The raw score for the language development survey is the total number of circled words. The manual provides

instructions for converting raw scores into T-scores.

Interpretability: Although the ASEBA provides the user with T-scores to compare a child's performance against other children and the scoring forms classify scores as normal (under 93 percent), borderline (93 to 97 percent), or clinical (over 97 percent), the authors recommend that the results be interpreted by someone with some graduate training.

Adaptations/Special Instructions for Individuals with Disabilities: The manual does not provide details about this, but suggests that persons rating children with disabilities compare

the child's behavior to their expectations of a typical same-age child.

Training Support: None indicated, however, ASEBA was designed to be easy to use and some support may be available on the internet.

Report Preparation Support: The manual shows a report generated by computer scoring software.

References:

Achenbach, Thomas M. and Leslie A. Rescorla. *Manual for the ASEBA Preschool Forms and Profiles*. ASEBA, Burlington, VT, 2000.

AGES & STAGES QUESTIONNAIRES, THIRD EDITION (ASQ-3), 2009 (2011 Update)

Authors:

Jane Squires, Elizabeth Twombly, Diane Bricker, and LaWanda Potter

Publisher:

Paul H. Brookes Publishing Co.
(800) 638-3775
<http://www.pbrookes.com>

Instrument:

<http://www.brookespublishing.com/resource-center/screening-and-assessment/asq/asq-3/>
<http://agesandstages.com/>

Initial Material Cost:

ASQ-3 Starter Kit (includes 21 photocopiable print masters of the questionnaires and scoring sheets, CD-ROM with printable PDF questionnaires, User's Guide, Quick Start Guide): \$249.95

Spanish Starter Kit with an English User's Guide: \$249.95

Representativeness of Norming Sample:

The sample represents English-speaking families from all 50 states and some U.S. territories and consists of 15,138 children (47.4 percent female) between 1 and 66 months of age. Children from both risk and non-risk populations are included. Families are educationally and economically diverse, and their ethnicities roughly match estimates from the 2007 U.S. Census. Data were collected between January 2004 and June 2008. The Spanish-speaking sample is described below (see Other Languages).

Languages:

English, Spanish

Type of Assessment:

Parent report

Age Range and Administration Interval:

1 to 66 months of age; 21 age-specific questionnaires for use at 2, 4, 6, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 27, 30, 33, 36, 42, 48, 54, and 60 months of age. Users may vary the interval to fit their needs.

Personnel, Training, Administration, and Scoring

Requirements: Questionnaires are written at no higher than a grade 6 reading level so that parents may easily understand and administer. Each questionnaire takes 10 to 15 minutes to administer and approximately 1 to 5 minutes to score. Scoring and interpretation requires professionals or trained paraprofessionals.

Brookes On Location, the professional development program of Brookes Publishing, offers onsite training for the ASQ-3 and ASQ:SE. Introduction seminars last 1 to 2 days and cost \$1,550 to \$2,800 for up to 40 attendees. Three-day train-the-trainer session costs include an individual registration fee (\$895), User's Guides (\$50 each), and lodging. Discounted fees may be available for group registrations of four or more.

DVDs are available on using the ASQ-3 system in the context of a program home visit (\$49.95) and on scoring and interpreting questionnaires (\$49.95).

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 3 (mostly .65 or higher)

Validity: 3 (concurrent, given in percentage, not correlation); 1 (predictive, none described)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 2 (self-administered by parent and scored by trained staff member or parent)

Description: The Ages & Stages Questionnaires, 3rd Edition (ASQ-3) is a series of 21 parent-completed questionnaires to help screen infants and young children for developmental

delays during their first 5.5 years. It is completed by parents or caregivers for children 1 to 66 months of age. Each questionnaire includes 30 developmental items and focuses on assessment of

five key developmental areas: Communication, Gross Motor, Fine Motor, Problem Solving, and Personal-Social. Communication addresses babbling, vocalizing, listening, and understanding; Gross Motor focuses on body movements, including arm, body, and leg; Fine Motor addresses finger movements; Problem Solving focuses on learning and playing with toys; and Personal-Social addresses social play and play with toys and other children. Parents rate each item as “Yes” the child does the behavior, “Sometimes,” and “Not Yet.” Items about behaviors that are challenging to describe (for example, putting beads on a string) include illustrations to help parents guide their responses. The items include a mix of skill ranges, including activities the child may not have tried before. Usually program or clinical staff work with parents to complete the questionnaires and encourage the parent to assist the child in trying an activity to see if they can do it (for example, if the child has not had the opportunity to draw a line, the parent or staff member might give the child a crayon and paper to see if she makes a line). The ASQ-3 also includes a section for parents to record general concerns and issues not captured by the questionnaire. One reviewer notes that the measure is a valuable screening tool for identifying potential developmental disabilities that are suspected in infants and young children but that it is not a standardized measure to be used for placement decisions for special education services (Hanig 2010).

The ASQ-3 updates the second edition of the ASQ (Squires and Bricker 1999) to include a new

standardization sample, new questionnaires administered to 2- and 9-month-olds, a new monitoring zone range to identify infants and children at risk of developmental delays but not scoring below cut-off points, open-ended questions, revised cut-off points, an updated User's Guide to facilitate use with diverse populations, and expanded administration windows. The authors and their colleagues are developing a version of the ASQ for children ages 4 weeks to 39 months, called the ASQ: Inventory, designed to monitor children's development using one form. This would provide a way to scale the items and obtain scores that are more comparable to a traditional assessment with basals and ceilings.

Other Languages: The Spanish questionnaires were field tested with Spanish-speaking parents in a number of states across the United States, including Arizona, Texas, and Washington, although authors do not report when the testing was conducted. Selected experts who work with Spanish-speaking children and families reviewed the second edition of the ASQ for accuracy and to determine if it would accommodate the variety of Spanish dialects spoken by families in the United States (for example, Nicaraguan, Mexican, Argentinean). The panel corrected translation errors and changed wording based on its review.

Although separate cut-off points for the Spanish questionnaires were not empirically derived as for the English questionnaires, cut-off points are similar for the Spanish and English samples at risk for developmental delays.

Uses of Information: The ASQ-3 provides comprehensive initial screening for developmental delays, monitoring and identification of areas in need of further assessment, and parent education about child development. It also fosters parent involvement.

Reliability: (1) Internal consistency: Cronbach's alpha was reported for each developmental area among children 2 to 36 months and 42 to 60 months, respectively: Communication (.57 to .80 and .66 to .83); Gross Motor (.57 to .87 and .68 to .73); Fine Motor (.51 to .77 and .76 to .83); Problem Solving (.53 to .78 and .70 to .78); and Personal-Social (.51 to .67 and .66 to .71).

(2) Test-retest reliability: percentage agreement between administrations within a two-week interval for 145 parents was 92 percent. Intraclass correlations ranged from .75 to .82.

(3) Inter-rater reliability: percentage agreement between parents and trained examiners was 93 percent. Intraclass correlations by developmental area ranged from .43 (Communication) to .69 (Personal-Social).

Validity: (1) Content validity: revisions were made to items included in the third edition based on ASQ user feedback and statistical analyses, including item response theory (IRT) modeling. Items that did not fit a developmental model underwent review and revision. Revisions included item wording adjustments, clarifications to examples accompanying items, modification of item illustrations, and the addition of an item to

the Communication developmental area to assist in identifying children who may have delays in expressive language. Items included in the Spanish version of the ASQ-3 underwent revision to correct translation errors in the second edition. Items were also reworded to mirror changes in items and format made in the ASQ-3 English version.

(2) Concurrent validity: a sample of 579 children was administered the ASQ-3 and Batelle Developmental Inventory (BDI) for classifications of typical development. Sensitivity between the ASQ-3 and BDI across all age groups (2 to 60 months) and developmental areas was 86 percent and ranged from 85 percent (2- to 12-month questionnaires) to 89 percent (14- to 24-month questionnaire) for infants and toddlers. Sensitivity for older children 42 to 60 months was 83 percent. Specificity between the ASQ-3 and BDI across all age groups and developmental areas was 86 percent and ranged from 78 percent (14- to 24-month questionnaires) to 91 percent (2- to 12-month questionnaires) for infants and toddlers and was 92 percent (42- to 60-month questionnaire) for older children.

(3) Construct validity: the developmental area scores are strongly correlated with the overall ASQ-3 scores, with the coefficients ranging from .65 to .70. The correlations among the developmental area scores in the moderate range (.33 to .54).

Bias Analysis: Parents and caregivers from the normative sample completed either paper questionnaires (n = 9,733) or web-based

questionnaires (n = 8,839). Statistical analyses, including IRT modeling, tested variations between the two types of questionnaires. Results indicated that 60 items out of 570 (10 percent) showed differential item functioning (DIF) and were equally distributed among all age intervals.¹ Within each developmental area, significant t-values (which is a statistical test for DIF) for items that showed DIF were not all positive or negative, indicating no consistent pattern of differences between the same items on the paper and web-based questionnaires.

Methods of Scoring: Parents provide one of three responses for each item in the ASQ-3 (i.e., yes, sometimes, not yet). Trained program staff or professionals convert item responses to point values and sum them for each developmental area.

Interpretability: Scorers transfer identifying information, item responses, and scores for each developmental area to an Information Summary Sheet. The sheet provides age-based normative score ranges, to which developmental area scores are compared. Scores are then determined as above, close to, or below cut-off points. Professionals or paraprofessionals are required to provide feedback to parents who have completed the questionnaire. They may provide parents with immediate feedback if scoring is conducted on site.

Training Support: The User's Guide contains complete instructions for each phase of the

questionnaire. Brookes On Location, the professional development program of Brookes Publishing, offers onsite one-day introductory seminars for the ASQ-3 and/or ASQ:SE as well as three-day train-the-trainer seminars. Two DVDs provide guidance on using the ASQ-3 system in the context of a home visit and on scoring and interpreting questionnaires. Other support materials include guidelines for choosing referral criteria to determine if children need more extensive assessment or close monitoring or little or no monitoring as well as activities sheets with games and events that correspond to the ASQ-3 age intervals.

Adaptations/Special Instructions for Individuals with Disabilities: No information available.

Report Preparation Support: Program staff may retain Information Summary Sheets as a record of the child's performance on the ASQ-3 so that questionnaires may be returned to parents for reference.

¹ This analysis used the 19 intervals from 4 to 60 months. The 2- and 9-month intervals were still under development.

References:

- Brookes Publishing Co. "Ages & Stages Questionnaires, Third Edition (ASQ-3)." Available at [<http://www.agesandstages.com>]. 2009.
- Farrell, Jane, LaWanda Potter, and Arden Munkres. *Ages and Stages Questionnaires on a Home Visit DVD*. Baltimore: Brookes Publishing Co., 2007.
- Hanig, Kenneth M. "Review of the Ages & Stages Questionnaire: A Parent-Completed Child Monitoring System, Third Edition." In *The Eighteenth Mental Measurements Yearbook*, edited by Robert A. Spies, Janet F. Carlson, and Kurt F. Geisinger. Lincoln, NE: The Buros Institute of Mental Measurements, 2010.
- Squires, Jane, and Diane Bricker. "Ages & Stages Questionnaires (ASQ): A Parent-Completed, Child-Monitoring System, Second Edition." Baltimore: Brookes Publishing Co., 1999.
- Squires, Jane, Diane Bricker, Elizabeth Twombly, Robert Nickel, Jantina Clifford, Kimberly Murphy, Robert Hoselton, LaWanda Potter, Linda Mounts, and Jane Farrell. *Ages & Stages English Questionnaires, Third Edition (ASQ-3): A Parent-Completed, Child-Monitoring System*. Baltimore: Paul H. Brookes Publishing Co., 2009.
- Squires, Jane, Diane Bricker, Elizabeth Twombly, Robert Nickel, Jantina Clifford, Kimberly Murphy, Robert Hoselton, LaWanda Potter, Linda Mounts, and Jane Farrell. *Ages & Stages Spanish Questionnaires, Third Edition (ASQ-3): A Parent-Completed, Child-Monitoring System*. Baltimore: Paul H. Brookes Publishing Co., 2009.
- Squires, Jane, Elizabeth Twombly, Diane Bricker, and LaWanda Potter. *ASQ-3 User's Guide, Third Edition*. Baltimore: Paul H. Brookes Publishing Co., 2009.
- Twombly, Elizabeth, and Jane Squires. *ASQ-3 Scoring and Referral DVD*. Baltimore: Brookes Publishing Co., 2009.

AGES AND STAGES QUESTIONNAIRES: SOCIAL-EMOTIONAL (ASQ:SE) A PARENT-COMPLETED, CHILD-MONITORING SYSTEM FOR SOCIAL-EMOTIONAL BEHAVIORS, 2002

Authors:

Jane Squires, Diane Bricker, and Elizabeth Twombly

Publisher:

Paul H. Brookes Publishing Co.

1-800-638-3775

www.brookespublishing.com

Initial Material Cost:

Questionnaires and User's Guide: \$125

Representativeness of Norming Sample:

Compared with 2000 Census figures, the normative sample underrepresents Caucasians and overrepresents individuals of mixed ethnicity and has higher percentage of well-educated mothers and low-income families.

Languages:

English and Spanish

Type of Assessment:

Parent report

Age Range and Administration Interval: 3 to 66 months; administered within 3 months of the target ages of 6, 12, 18, 24, and 30 months, and within 6 months of the target ages of 36, 48, and 60 months

Personnel, Training, Administration, and Scoring

Requirements: The ASQ:SE can be administered by parents, child care providers, and preschool teachers (10 to 15 minutes per questionnaire). Ideally, program staff will train parents on administering the ASQ:SE; training takes approximately 2 to 3 hours. Scoring should be done by a paraprofessional, and should only take a few minutes per questionnaire.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 3 (.50 or higher)

Norming Sample Characteristics: 2 (not nationally representative, normed within the past 15 years)

Ease of Administration and Scoring: 3 (self-administered by parent but scoring by paraprofessional recommended)

Description: This series of eight parent-completed questionnaires with 22 to 36 items in each questionnaire helps determine children's progress in their social-emotional behavior. Each questionnaire can be used within 3 months of the target age (for the 6- through 30-month questionnaires) or 6 months of the target age (for the 36- through 60-month questionnaires). The questionnaires focus on seven behavioral areas: (1) self-regulation, (2) compliance, (3) communication, (4) adaptive functioning, (5) autonomy, (6) affect, and (7) interaction with people. Each questionnaire is written at a 5th- to 6th-grade reading level. The ASQ:SE can be used to screen for social-emotional development

problems at one point in time or to monitor a child repeatedly at different intervals. The publisher recommends that the ASQ:SE be used in conjunction with a developmental screening tool that provides information on the child's communication, motor, and cognitive functioning.

Uses of Information: The ASQ:SE was developed to complement the ASQ by providing information specifically addressing the social and emotional behavior of children ranging in age from 3 to 66 months. The ASQ:SE is a screening tool that helps practitioners identify infants and young children whose social or emotional

development requires further evaluation to determine whether referral for intervention services is necessary.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): the alphas for the questionnaires were .82 overall, .69 (6-month), .67 (12-month), 18-month (.81); 24-month (.80); 30-month (.88); 36-month (.89); 48-month (.91); 60-month (.91). (2) Test-retest reliability, with one to three weeks between tests: percent agreement between scores by the same rater on two occasions is 94 percent. (3) Inter-rater reliability: no information available.

Validity: (1) Concurrent validity: percent agreement of ASQ:SE with similar established tools ranged from 81 to 95 percent and was 93 percent overall. (2) Predictive validity: no information available.

Method of Scoring: Scoring can be done by paraprofessional or professional staff. Scoring options for the items in the ASQ:SE are "most of the time," "sometimes," and "rarely or never." Each response is converted to a numerical value. The numerical values are totaled and compared with the empirically derived cutoff score (for that particular questionnaire interval) that indicates whether a child should receive further in-depth evaluation. The reproducible scoring sheets all include referral considerations that help determine whether the child needs further evaluation.

Interpretability: An Information Summary Sheet is provided to assist program staff with

scoring and summarizing assessment information and providing them with a summary of the child's performance on the questionnaire. The Information Summary Sheet contains instructions for scoring the questionnaire, a chart indicating cutoff scores for referrals, and a list of considerations prior to making referrals to mental health professionals. Children whose scores are at or greater than the cutoff point should be considered for further evaluation or referral, and children with scores below the cutoff point can be monitored with another ASQ:SE in 6 to 12 months.

Training Support: The User's Guide contains complete instructions for training on the ASQ:SE, setting up the assessment, and conducting it. It provides instructions for administering the questionnaires with sensitivity to children's environmental, cultural, and social-emotional differences. Other support materials include compilation of detailed technical data on how the system was developed and tested, case examples, and creative activities and lists of social-emotional behaviors professionals can share with parents for use with each age group. The publisher also offers customized training seminars to provide guidance on using this assessment tool.

Adaptations/Special Instructions for Individuals with Disabilities: The ASQ:SE User's Guide briefly mentions the importance of interpreting assessment information within the context of the specific child's health, development, and family/cultural factors. The guide also

describes factors to consider before making a referral based on the ASQ:SE assessment.

Report Preparation Support: None described.

References:

Squires, Jane, Diane Bricker, and Elizabeth Twombly. *Ages And Stages Questionnaires: Social-Emotional (ASQ:SE), A Parent-Completed, Child-Monitoring System For Social-Emotional Behaviors*. Baltimore: Paul H. Brookes Publishing Co., 2002.

Squires, Jane, Diane Bricker, and Elizabeth Twombly. *Ages And Stages Questionnaires: Social-Emotional (ASQ:SE), A Parent-Completed, Child-Monitoring System For Social-Emotional Behaviors, User's Guide*. Baltimore: Paul H. Brookes Publishing Co., 2002.

For information on continuing research and adjustments in cutoff points on the ASQ:SE, refer to <http://www.brookespublishing.com/resource-center/screening-and-assessment/asq/asq-se/>.

ASSESSMENT, EVALUATION, AND PROGRAMMING SYSTEM (AEPS) MEASUREMENT FOR BIRTH TO THREE YEARS, 1993

Authors:

Diane Bricker, Juliann Cripe, Kristine Slentz

Publisher:

Paul H. Brookes Publishing Co.

1-800-638-3775

www.brookespublishing.com

Initial Material Cost:

AEPS™ Birth to Three set (Administration Guide, Test, Curriculum for Birth to Three Years): \$150

Representativeness of Norming Sample: No norming sample.

Languages:

English¹

Type of Assessment:

Observation, direct assessment (to elicit a behavior), and parent, caregiver, or therapist report

Age Range and Administration Interval: Children whose developmental age is 3 years or less. May be used for children whose chronological age is 6 years or less. Readministered at 3- or 4-month intervals.

Personnel, Training, Administration, and Scoring Requirements: The AEPS can be used by both direct service personnel and specialists. Administration time may range from 1 to 2 hours for the initial assessment and 15 to 30 minutes for subsequent assessments, depending on the child's level of functioning and the user's familiarity with the AESP and the child. Quarterly or yearly followups generally take half the time of the initial assessment.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher).

Validity: 3 (.5 or higher for content validity) Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (administered and scored by someone with basic clerical skills)

¹ The first edition also has Spanish translations of the Family Interest Survey and the Family Report.

Description: The Assessment, Evaluation, and Programming System (AEPS) Measurement for Birth to Three Years is a criterion-referenced assessment tool that is designed to help early interventionists improve their assessments of the abilities and needs of young children who have disabilities or are at risk for developmental delays. The AEPS test was designed to be used in conjunction with the AEPS Curriculum for Birth to Three Years or other similar curricula. It covers the developmental progress of children's functional skills in six key domains (fine motor, gross motor, adaptive, cognitive, social-

communication, and social development). Each domain is divided into strands, which consist of related groups of behavior divided into common categories. Each strand has a series of goals and discrete objectives that lead up to the goal. The strands, goals, and objectives are developmentally sequenced. Objectives and goals are either observed, elicited, or recorded based on parent, caregiver, or therapist report.

The AEPS encourages family participation in the assessment through the use of family-focused materials, such as the family report, planning

guide, child progress record, and family interest survey.

Uses of Information: The AEPS is an assessment/evaluation tool that is used to create individual evaluation programs and intervention plans, known as Individualized Educational Program/Individualized Family Service Plans (IEP/IFSP). The test is used to provide a baseline on the child's functioning to help develop the intervention curriculum and to measure the child's developmental progress over time.

Reliability: Several research samples were drawn from children ages 2 months to 6 years in early intervention programs in Oregon, Washington, Idaho, and Arkansas, and 48 interventionists in Vermont, Iowa, Oregon, and British Columbia. (1) Inter-rater reliability: Pearson product moment correlations for individual domains ranged from .71 for the Social Domain to .96 for the Gross Motor Domain. Mean correlation for all domains was .88. Total test score correlation was .97. (2) Test-retest reliability (1-2 week interval): Pearson Product Moment correlation for domains ranged from .77 for the Social Domain to .95 for the Gross Motor Domain, with a correlation of .88 for all domains. Total test agreement was .95.

Validity: (1) Congruent Validity: Pearson Product Moment correlations with the Bayley Scales of Infant Development Mental Age and Motor Age were .93 and .88, respectively. Correlation with the Gesell Developmental Scale Maturity Age scores was .51.

Method of Scoring: Each of the six domains has a specific recording form. Items are marked as "pass consistently" (2), "inconsistent performance" (1), and "does not pass" (0). Specific criteria are provided for each goal and objective. In addition to scoring each of the items, a qualifying note is attached to each item goal and objective. Items are marked as "assistance provided" (A), "behavior interfered" (B), "reported assessment" (R), "modification/adaptation" (M), and "direct test" (D).

Scoring can be done two ways: (1) a total score is computed for domains by counting the number of goals and objectives scored with a "2" in each domain. For the total frequency, the domain scores are added together. The number of "1" scores are computed in the same way. (2) The percentage of items scored with a "2" or "1" can also be calculated by dividing the total "2" scores by the total number of items in the domain and the total "1" scores by the total number of items in the domain. For an overall percent score, the total number of items scored with a "2" across the domains is multiplied by 2 and divided by 456 (total number possible).

The AEPS Family Interest Survey and AEPS Family Report, which are family-centered materials that can be used in conjunction with the AEPS Test, are completed by families and have separate scoring guidelines.

Interpretability: No instructions provided. However, the scoring of the instrument will inform the interventionist how well the child is

performing the observed skill. The test results expand the quantity and quality of developmental information and help professionals develop IFSP or IEP goals and objectives.

Training Support: “Brookes on Location” professional development seminar, AEPS™: A Linked System of Assessment, Intervention, and Evaluation, is available through the publisher (www.brookespublishing.com/).

Adaptations/Special Instructions for Individuals with Disabilities: The AEPS was designed for use with populations of children who are at risk and who have disabilities. For children who have severe impairments, general modifications are required. For children with severe disabilities, the AEPS test objectives should be used more as goals, and the associated

curricular programming steps as objectives. General adaptation guidelines are provided for children with visual, hearing, and motor impairments.

Report Preparation Support: Data Recording Forms are provided for scoring tests, graphing results, tracking scores, and recording comments. A Child Progress Record is available to track progress on strand objectives.

References:

Cripe, Juliann, Kristine Slentz, and Diane Bricker. AEPS Curriculum for Birth to Three Years, Volume 2. Baltimore, MD: Paul H. Brookes Publishing Co., Inc., 1993.

BATTELLE DEVELOPMENTAL INVENTORY (BDI), 1984

Authors:

J. Newborg, J.R. Stock, & J. Wnek (initial development); J. Guidubaldi (pilot norming study); J.S. Svinicki (completion and standardization)

Publisher:

Riverside Publishing Co.
800-323-9540
www.riverpub.com

Initial Material Cost:

Examiner's manual: \$58
Scoring booklets (15): \$35
Examiner's manual for screening test: \$56
Screening test booklets (30): \$41
Overview videotape: \$58

Representativeness of Norming Sample: National norming sample of 800 children from birth to 8 years of age, stratified according to geographical region, age, race, and gender; 75 percent urban and 25 percent rural; 28 test sites in 24 states. Distribution of sample closely represents the four major geographical regions of the United States.

Languages:

English

Type of Assessment: Direct Child Assessment, Observation, and Parent Interview

Age Range and Administration Interval: Birth to 8 years.

Personnel, Training, Administration, and Scoring

Requirements: Can be administered by paraprofessionals ("nonpsychologists") and is intended for use by infant, preschool, primary, and special education teachers. Important that examiners have supervised practice in administering BDI for children with disabilities across age span.

BDI Screening Test takes 10 to 15 minutes for children under 3 and over 5 years of age and 20 to 30 minutes for children between the ages of 3 and 5. The full BDI can be administered in about 1 hour for children under 3 and over 5 years of age and in 1.5 to 2 hours for children between 3 and 5 years.

Summary:

Initial Material Cost: 1 (< \$100)
Reliability: 3 (test-retest, .65 or higher)
Validity: 3 (.5 or higher for concurrent)
Norming Sample Characteristics: 2 (older than 15 years, nationally representative)
Ease of Administration and Scoring: 2 (administered and scored by paraprofessionals)

Description: The BDI assesses children from birth to 8 years of age on the following five domains: Personal-Social, Adaptive, Motor, Communication, and Cognitive. The BDI Screening test contains 96 items and represents a subset of the full battery, which is comprised of 341 items. Within each domain, the items are assigned age levels and organized sequentially into subdomains. The BDI is norm-referenced and helps to identify young children with special needs

and assess the functional abilities of these children, as well as children without special needs. Child diagnostic information for the full BDI is presented in the form of age equivalents, percentiles, and standard scores (that is, developmental quotients, z-scores, T-scores, and normal curve equivalents) for each of the major domains and subdomains. This information is available for the Screening Test as age equivalents and cutoff scores.

Uses of Information: The BDI is primarily used for four purposes: (1) assessment and identification of children with special needs, (2) assessment of school readiness among children without special needs, (3) planning and providing educational instruction, and (4) evaluation of groups of children with special needs.

Reliability: Test-retest (4-month interval) for the total test is .98 for children 0 to 5 and 18 to 23 months old and .99 for children 6 to 11, 12 to 17, 24 to 35, and 36 to 47 months old.

Validity: (1) Content validity: The process for developing the BDI involved identifying skill areas to be assessed, selecting or developing the test items, and verifying the content validity of the results with review by content experts. (2) Construct validity: Factor analysis and the intercorrelations between the domains and subdomains supported the factorial validity and conceptual structure of the BDI. (3) Concurrent validity: Measures on the BDI relate well to other instruments, including the Vineland Social Maturity Scale (Doll 1965), and the Developmental Activities Screening Inventory (DASI; Dubose & Langley 1977), with correlations ranging from .78 to .94. While the BDI is not an intelligence test, it measures motor and language skills and is found to relate moderately well with the Stanford-Binet Intelligence Scale (S-B; Terman & Merrill 1960), with correlations ranging from .40 to .61. Validity tests were also conducted with the BDI Screening Test. The correlation between the total score on the Screening Test and that of the full battery is .99, indicating that performance

on the Screening Test predicts performance on the full BDI.

Method of Scoring: Items are scored on a three-point system, according to whether the child typically completes the item correctly (2), sometimes does so (1), or rarely or never completes the task, even if the child did not have the opportunity to respond (0). The number of allowed trials is presented separately with each item. If a child completes the item correctly on the first trial, he or she receives 2 points and can move on to the next item. Basal rules are established so that test items that are extremely easy for a child need not be administered, while ceiling rules ensure that items that are much too difficult are not administered. A child receives full credit, 2 points per item, for all items that fall below the basal level. The subdomain raw scores for the full BDI battery are obtained by summing the individual item scores from the basal level through the ceiling level and then adding that total to the full credit sum (2 points per item) for items below the basal level. In contrast, for the Screening Test the basal and ceiling rules apply to each domain. Domain raw scores can be obtained by summing the subdomain raw scores, and a total raw score is obtained by summing the five domain raw scores. Once obtained, the raw scores are transferred to the Score Summary and Profile section in the Scoring Booklet.

Interpretability: Tables are used to convert raw scores to percentile rank, age equivalent scores, and the following standard scores: z-scores, T scores, deviation quotients, and normal

curve equivalents. The Screening Test cutoff scores for each age group are provided for three probability levels that correspond to 1.0, 1.5, and 2.0 standard deviations below the mean. In interpreting the full BDI, standard scores are useful for decision making, percentile ranks are useful for reporting information to parents, and age equivalent scores may be required by federal, state, and local policies. Cutoffs are not provided for the full BDI, but the authors recommend following convention and treating standard scores that are 1.5 or more standard deviations below the mean as an indication of a performance deficit. BDI norms should not be used if the BDI has not been administered according to standard procedures. Five case studies on the interpretation of scores are presented in chapter 4 of the manual.

Training Support: An overview videotape is available.

Adaptations/Special Instructions for Individuals with Disabilities: The BDI includes guidelines for assessing children with disabilities so that they are able to respond in a manner that is appropriate for them. Also, most of the items include standardized stimulus/response options for children with visual, hearing, neuromotor, or behavior/emotional needs. Children with special needs are scored according to the same criteria used to score children without disabilities. Adaptations are not made for children who have

“no opportunity” for a response on an item due to handicapping conditions, because a score of 0 reflects children’s actual level of functioning.

Report Preparation Support: Guidelines for developing goals and objectives for children with special needs through the Individual Education Plans are found in chapter 5 of the manual.

References:

Doll, E.A. Vineland Social Maturity Scale. Circle Pines, MN: American Guidance Service, 1965.

DuBose, R.F. and M.B. Langley.

Developmental Activities Screening Inventory. Hingham, MA: Teaching Resources Corporation, 1977.

Newborg, Jean, John Stock, Linda Wnek, John Guidubaldi, and John Svinicki. Battelle Developmental Inventory with Recalibrated Technical Data and Norms: Examiner’s Manual. Riverside Publishing, 1984.

Newborg, Jean, John Stock, Linda Wnek, John Guidubaldi, and John Svinicki. Battelle Developmental Inventory with Calibrated Technical Data and Norms: Screening Test. Riverside Publishing, 1984.

Terman, L.M. and M.A. Merrill. Stanford-Binet Intelligence Scale, Third Revision (Form L-M). Boston, MA: Houghton Mifflin, 1960.

BAYLEY SCALES OF INFANT AND TODDLER DEVELOPMENT, THIRD EDITION (BAYLEY-III), 2005 (2011 Update)

Authors:

Nancy Bayley

Publisher:

Pearson

(800) 627-7271

<http://www.pearsonassessments.com>

Instrument:

<http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8027-23X&Mode=summary>

Initial Material Cost: Complete Kit (includes Administration and Technical Manuals; 25 each of Cognitive, Language Composite, and Motor Composite Record Forms; Stimulus Book; Picture Book; Manipulative Set; 25 Social-Emotional and Adaptive Behavior Questionnaires; 25 Caregiver Report Forms; and Rolling Case): \$950

Representativeness of Norming Sample¹: The norming sample for the Cognitive, Language Composite, and Motor Composite scales was a national, stratified sample of 1,700 children 1 to 42 months old (year not

specified). The sample included 100 children in each of 17 age groups (age group spanned between 1 and 4 month intervals). Based on the 2000 U.S. Census Bureau's Current Population Survey, the sample was stratified along the following variables: gender, region, race/ethnicity, and parent education. The children were recruited from health clinics, hospitals, child development centers, churches, and other community organizations and identified by professional recruiters. The initial sample was restricted to typically developing children; then, a subgroup of children with special needs who participated in test development trials (about 10 percent of the sample) was included.

Languages:

English

Type of Assessment: Direct child assessment, observation, and caregiver report

Age Range and Administration Interval: 1 to 42 months old. Can be administered at 3-month intervals for infants less than 12 months of age and at 6-month intervals for older children.

Personnel, Training, Administration, and Scoring Requirements: Administered by a highly trained program staff member. Trained individuals without graduate or professional training in assessment may administer and score the assessment under supervision. Average administration time is 50 minutes for infants up to 13 months old and 90 minutes for children 13 months old and older.

A training DVD and an interactive administration and scoring DVD are available for \$61.20 and \$90.75, respectively.

Summary:

Initial Material Cost: 3 (> \$200)

Reliability: 3 (mostly .65 or higher)

Validity: 3 (mostly .5 or higher for concurrent)

Norming Sample Characteristics: 3 (normed within the past 15 years, nationally)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

¹ The Social-Emotional Scale was normed during the Bayley-III test development phase in spring 2003. The norming sample was a sample of 456 U.S. children 1 to 42 months old stratified by region, race/ethnicity, and parent education. The sample included eight age groups of approximately equal numbers of males and females, with 50 to 89 children per group. Children were excluded from the sample if they did not speak or understand English; had hearing or visual impairments; had developmental risk factors based on social, socioeconomic status, or parent education factors; or were taking medication that could affect test performance. The Adaptive Behavior Scale standardization, reliability, and validity evidence is based on the norming sample data of the Adaptive Behavior Assessment System-Second Edition, which included a stratified sample of 1,350 children 0 to 71 months old divided into 13 groups of 3- to 5-month intervals of 100 children each, except for the oldest age group, which had 150 children. Each group included equal numbers of males and females. The sample was stratified by race/ethnicity and parent education level, and efforts were made to ensure that the sample was geographically representative. Children with special needs represented 2.88 percent of the sample.

Description: The Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III) is an individually administered assessment that measures the developmental functioning of infants and children 1 to 42 months old. The measure presents children with situations and tasks designed to produce an observable set of behavioral responses that are assessed directly on the following scales: Cognitive Scale (91 items), Language Composite Scale (97 items) with Receptive and Expressive Language subscales, and Motor Composite Scale (138 items) with Fine- and Gross-Motor subscales. The child's parent or primary caregiver completes two additional scales: Social-Emotional and Adaptive Behavior. The assessor completes a Behavior Observation Inventory at the conclusion of the assessment to determine how often behaviors, such as positive affect and cooperativeness, are observed during testing. The assessor asks the child's caregiver to rate the degree to which the child typically exhibits the behavior. Though not discussed here, a 15- to 25-minute screener is also available to assess cognitive, language, and motor development (Bayley-III Screening Test) for use in ongoing developmental screening.

The Bayley-III updates the Bayley Scale for Infant Development-Second Edition (BSID-II), published in 1993, and divides the Mental Scale into Cognitive and Language Composite scales. The BSID-II Behavior Rating Scale was revised and replaced with the Social-Emotional and Adaptive Behavior scales. The Bayley-III adds new items to the Cognitive, Language Composite, and Motor Composite scales and includes the

Behavior Observation Inventory to gauge whether a child's behavior during the assessment was representative of the child's typical conduct. The updated measure extends the floor and ceiling for each scale by including gifted children and children with or at risk of developmental challenges. Stimulus materials were updated and printed in color, and procedures were modified to increase ease of administration. In addition, the Bayley Short Form-Research Edition (BSF-R; Andreassen and Fletcher 2005), based on the BSID-II, was developed for the Early Childhood Longitudinal Study-Birth Cohort to provide an assessment that is less time-consuming than the full BSID-II but sufficiently comprehensive to capture adequately the development of infants and young children (West and Andreassen 2002).

The Social-Emotional Scale comprises items assessing social-emotional competence and sensory processing; it is based on the Greenspan Social-Emotional Growth Chart: A Screening Questionnaire for Infants and Young Children (Greenspan 2004). The assessment measures functional emotional milestones. The Adaptive Behavior Scale, based on the Adaptive Behavior Assessment System-Second Edition (ABAS-II; Harrison 2003), assesses the attainment of adaptive behavior skills necessary for infants' and young children's development of independence.

Other Languages: None.

Uses of Information: The Bayley-III is used to identify areas of relative impairment or delay, develop curricula for interventions, and assess the outcome of such interventions. It is not a

diagnostic tool but rather indicates areas that might require further evaluation. The scales should not be used to assess a child's deficit in a specific skill area or to make a norm-referenced interpretation of scores for children with severe sensory or physical impairments. In addition, although some of the measure's items are similar to items on tests of school-age abilities, the Bayley-III is not an intelligence test.

Reliability: (1) Internal consistency reliability: the author provided split-half reliability coefficients for the Cognitive, Language Composite, and Motor Composite scales on the entire norming sample. The coefficients for scores on the Cognitive Scale ranged from .79 to .97 across the 17 age groups. The coefficients for scores on the Language Composite Scale ranged from .82 to .98 and, across its subscales, from .71 to .97. For raw scores on the Motor Composite Scale, the coefficients ranged from .86 to .96 and, across its subscales, from .72 to .95.

The author also estimated the internal consistency of the Cognitive, Language Composite, and Motor Composite scales with a clinical population of 668 children. For scores on the Cognitive Scale, coefficients ranged from .90 to .99 across nine age groups. For the Language Composite Scale, reliability coefficients ranged from .74 to .99 across subscales. For the Motor Composite Scale, coefficients ranged from .84 to .99 across subscales.

For the Social-Emotional and Adaptive Behavior scales, the internal consistency estimates come from the original measures' manuals,

including the Greenspan Social-Emotional Growth Chart and ABAS-II. Reliability coefficients for the Greenspan Social-Emotional Growth Chart spanned across 8 age groups from 0 to 42 months. Coefficients ranged from .83 to .94 for the Social-Emotional items and from .76 to .91 for the Sensory Processing items. Reliability coefficients for the ABAS-II spanned across 10 age groups from 0 to 47 months. Coefficients ranged from .86 to .98 for the total score (i.e., General Adaptive Composite) across age groups. For subscale scores, they ranged from .70 to .96 for the Conceptual domain, from .81 to .94 for the Social domain, and from .82 to .97 for the Practical domain (except for the scores of children 0 to 3 months old at 0.65). For a sample of 246 children with developmental delays, motor impairments, language disorders, and biological risk factors, coefficients for the ABAS-II total score ranged from .97 to .99 while coefficients for the subscales ranged from .90 to .99.

(2) Test-retest reliability: test-retest reliability is based on scores from the Cognitive, Language Composite, and Motor Composite scales in 197 children 2 to 42 months old from the norming sample. The test-retest interval between administrations ranged from 2 to 15 days (mean = 6). The correlations for scores on the Cognitive Scale ranged from .75 to .86 across four age groups. The correlations for scores on the Language Composite Scale ranged from .69 to .87 and, for its subscales, from .63 to .84. The correlations for scores on the Motor Composite Scale ranged from .79 to .84 and, for its subscales, from .73 to .86. The author also presents

correlations corrected for the variability of the standardization sample. No test-retest reliability information is reported for scores of the Social-Emotional Scale. The author assessed test-retest reliability of scores from the Adaptive Behavior Scale by using parent reports for 207 infants and toddlers 0 to 35 months old. The test-retest interval between administrations ranged from 2 days to 5 weeks (mean = 12 days). The correlations for the total score ranged from .86 to .91 while correlations for subscale scores ranged from .81 to .90.

(3) Inter-rater reliability: inter-rater reliability was assessed for the Adaptive Behavior Scale for a sample of 56 children 0 to 71 months old, each rated by two parents. The correlation for the total scores was .77. Correlations ranged from .69 to .83 across domains.

Validity: (1) Content validity: Expert consultation and a review of the literature guided development of the Cognitive, Language Composite, and Motor Composite scales. In addition, an advisory panel lent guidance throughout the development process. The author also consulted clinical measurement specialists and conducted focus groups and surveys. Development of the Bayley-III occurred in several stages with pilot and tryout studies. The developers of the ABAS-II established its set of daily independent living skills based on legal and professional concepts, standards, and regulations related to special education and developmental disability. A confirmatory factor analysis using the norming sample data supported a three-factor

latent structure for the Cognitive, Language Composite, and Motor Composite scales (based on the root mean square error of approximation as the goodness-of-fit index).

(2) Concurrent validity:² During standardization, the author compared the Bayley-III scores to scores on the following instruments: the BSID-II, Weschler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III), Preschool Language Scale-Fourth Edition (PLS-4), and Peabody Developmental Motor Scales-Second Edition (PDMS-2). Separate samples were used for each combination of the Bayley-III with one other measure. Samples ranged from about 50 to 100 children 0 to 42 months old (except for the WPPSI-III, which included children 28 to 42 months old), and the test intervals ranged from 0 to 28 days (mean = 5 to 6). The administration order of the assessments was counterbalanced, with the means compared across the two orders of administration.

The correlations between scores on the Bayley-III Cognitive Scale and other cognitive measures were .60 for the BSID-II Mental Scale, .72 to .79 for the three WPPSI-III composite scales (Full, Verbal, and Performance IQs), and .57 for the PLS-4 Composite. Correlations between scores on the Bayley-III Language Composite Scale and other language measures were .71 for the BSID-II Mental Scale, .71 to .83 for the WPPSI-III composite scales, and .66 for the PLS-4 Composite with correlations of .62 and .68, respectively, with

² Correlations were corrected for the variability of the standardization sample.

the PLS-4 Receptive and Expressive communication subtest scores. The correlation between scores on the Bayley-III Motor Composite Scale and the PDMS-II Total Motor Quotient was .57 and, between the Bayley-III and PDMS-II fine- and gross-motor subtest scores, .59 and .57, respectively. The correlation between the Bayley-III Social-Emotional Scale and BSID-II Behavior Rating Scale was .38. The author reports that the ABAS-II (the basis for the Bayley-III Adaptive Behavior Scale) and Vineland Adaptive Behavior Scale-Interview Edition (VABS) were administered to a sample of 45 typically developing children 1 to 69 months old (mean = 34). The correlation between their scores on the two measures was .70.

Correlations of Bayley-III composites and subscales with assessments measuring largely different skills range in magnitude, with some at the moderate level. The correlations for the Bayley-III Cognitive Scale scores were .40 and .38, respectively, with the BSID-II Motor and Behavior Rating scales and .45 with the PDMS-II total motor score. The correlations between scores on the Bayley-III Language Composite Scale and the BSID-II Motor and Behavior Rating scales were .47 and .37, respectively, and .45 with the PDMS-II total motor score. The correlations between scores on the Bayley-III Motor Composite Scale and the BSID-II Mental and Behavior Rating scales were .44 and .31, respectively, .55 with the WPPSI-III, and .44 with the PLS-4 Composite. The correlations for the Bayley-III Social-Emotional Scale scores were .25 and .24, respectively, with the BSID-II Mental and Motor

scales scores, .43 with the WPPSI-III, .23 with the PLS-4 Composite, and .25 with the PDMS-II total motor score.

The author reports that the Bayley-III and ABAS-II were administered to a sample of 60 children 5 to 37 months old, with a test interval of 0 to 23 days (mean = 4). The correlations between scores on the Bayley-III scales and the ABAS-II General Adaptive Composite ranged from .25 to .36, except for the .04 correlation between the ABAS-II General Adaptive Composite score and the Social-Emotional scale. The ABAS-II and the Scales of Independent Behavior-Revised: Early Development Form (SIB-R), a brief 40-item assessment, were administered to a typically developing sample of 34 children 2 to 23 months old (mean = 14); the correlation between scores was .18.

The author assessed the Bayley-III's ability to differentiate (not diagnose) between convenience samples of special populations and a group of typically developing children matched on gender, parent education level, race/ethnicity, and geographic region. Compared to scores for children without disabilities, the scores on all subscales and composites were significantly lower for the following groups: children with Down syndrome, children with pervasive development disorder, children with cerebral palsy, children with specific or suspected language impairment, and children at risk for developmental delay. Children with asphyxiation at birth scored significantly lower than children in the matched control group on all subscale and composite

scores except for the Expressive Language subscale. Children with prenatal alcohol exposure had significantly lower scores on all subscales and composites except for the Motor Composite and its associated subscales. Children small for gestational age scored significantly lower on the Language Composite and Motor Composite and the Receptive Communication and Gross-Motor subscales. Children born prematurely or with low birth weight had significantly lower Motor Composite and Fine-Motor subscale scores.

The author also examined the Social-Emotional scale scores. The percentage of children scoring two or more standard deviations below the mean was greater for the special population groups (4 to 67 percent) than for the control groups (0 to 3 percent). In terms of the ABAS-II, the author found significant differences for all skill areas and domains between the following groups of children and their matched controls: children with developmental delays, children with biological risk factors, children with motor and physical impairments, and children with receptive and/or expressive language disorders.

(3) Predictive validity: analyses were conducted with the original Bayley Scales of Infant Development (BSID 1969), which included Mental and Motor scales. “. . . [T]he BSID at the scale level is generally not as predictive of later intellectual, language, or achievement performance as are specified subscales; however, the later in the preschool period (i.e., beyond two years) the BSID scores are obtained, the more

predictive they are of later childhood functioning” (Bayley 1993).

(4) Construct validity: intercorrelations between composite and subscale scores were .71 between the Language Composite Scale and each subscale score (Receptive and Expressive Communication) and .69 and .71 between the Motor Composite Scale and its subscale scores (Fine and Gross Motor). Higher correlations were observed between the Cognitive Scale and Language Composite subscales than between the Cognitive Scale and Motor Composite subscales. The author concluded that the relationship between language and cognition was closer.

Bias Analysis: During the test development process, potential item bias was assessed through expert review and statistical analyses, which led to the deletion of 30 items. The author tested an additional 120 black and Hispanic children to ensure adequate sample sizes for analyses of these groups. The analysis for differential item functioning (DIF) used the Mantel-Haenszel method, although the manual does not include the results.

Methods of Scoring: Assessors who are not highly trained may score the assessment under supervision. For the Cognitive, Language Composite, and Motor Composite scales, items are scored as correct (1) or incorrect (0) depending on whether the child displayed the indicated action. The administration manual provides detailed scoring instructions for each item. The raw score is the sum of the child's correct points. All items below the basal are scored

as correct, and all above the ceiling as incorrect. The Social-Emotional Scale uses a six-point frequency rating (can't tell, none of the time, some of the time, half of the time, most of the time, all of the time). The raw score is the sum of behavior frequencies. The total for the first eight items provides a sensory processing score. The Adaptive Behavior Scale uses a four-point frequency rating (is not able, never when needed, sometimes when needed, always when needed) and provides the following scores: a subscale score for each of the 10 skill areas; 3 domain area scores for the Conceptual, Social, and Practical domains; and a General Adaptive Composite score that is a sum of each child's scores across skill areas. The two ratings on the Behavior Observation Inventory by the assessor and caregiver are based on Likert-type scales for how often a behavior occurred during the observation (assessor rating) or how typical the behavior is (caregiver rating). While no scoring is completed, the qualitative comparisons are used with scores for intervention planning.

Using the tables provided in the manual, the assessor may convert raw scores into scaled scores (ranging from 1 to 19), composite scores, and percentile ranks and may determine confidence intervals. For the Cognitive, Language Composite, and Motor Composite scales, the assessor may plot scores on a chart to detect growth over time and to calculate developmental age equivalents for delayed children. Scoring software is available for purchase from the publisher.

Interpretability: Only persons with formal training in test administration should interpret

the results of the Bayley-III. The technical manual provides detailed information on interpretation. Norms are available by age groups of varying intervals (e.g., 10 days to 3 months) to facilitate norm-referenced interpretation of performance during the period of infant and toddler development. The Behavior Observation Inventory, completed by the assessor at the conclusion of the assessment, provides qualitative information to parents and caregivers to help them interpret the child's performance and to facilitate intervention planning.

Training Support: The Bayley-III may be purchased only by individuals highly trained in test administration and interpretation as evidenced by a doctoral degree, certification, or licensure. The administration manual provides detailed information on how to administer and score the assessment. A training DVD and an interactive administration and scoring DVD are available for purchase. A person with the above credentials may train and supervise others in administering and scoring the assessment.

Adaptations/Special Instructions for Individuals with Disabilities: The administration manual provides a detailed appendix of strategies for accommodating children with visual, hearing, and movement and motor impairments. For example, ambient noise should be reduced, and the assessor should speak clearly and naturally at an even pace for children with hearing impairments.

Report Preparation Support: Scoring Assistant Software is available for purchase to produce child

reports for caregivers to help them interpret results for individual children.

References:

Albers, Craig A., and Adam J. Grieve. "Test Review: Bayley Scales of Infant and Toddler Development-Third Edition." *Journal of Psychoeducational Assessment*, vol. 25, no. 2, 2007, pp. 180-190.

Andreassen, C., and P. Fletcher. "Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) Methodology Report for the Nine-Month Data Collection (2001-02), Volume 1: Psychometric Characteristics." Submitted to the U.S. Department of Education. Report No. (NCES 2005-100). Washington, DC: National Center for Education Statistics, 2005.

Bayley, Nancy. *Bayley Scales of Infant Development-Second Edition Manual*. San Antonio, TX: PsychCorp, 1993.

Bayley, Nancy. *Bayley Scales of Infant and Toddler Development-Third Edition: Administration Manual*. San Antonio, TX: PsychCorp, 2006.

Bayley, Nancy. *Bayley Scales of Infant and Toddler Development-Third Edition: Technical Manual*. San Antonio, TX: PsychCorp, 2006.

Greenspan, S.I. *Greenspan Social-Emotional Growth Chart: A Screening Questionnaire for Infants and Young Children*. San Antonio, TX: Harcourt Assessment, 2004.

Harrison, P.L., and T. Oakland. *Adaptive Behavior Assessment System, Second Edition*. San Antonio, TX: Psychcorp, Harcourt Assessment, 2003.

Pearson Assessments. "Bayley-III Enhanced Administration/Scoring Resource Interactive DVD." Available at [<http://pearsonassess.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8027-23X&Mode=detail&Leaf=accessory&dsrc=015-8027-612#ISBN2>]. 2005.

Pearson Assessments. "Bayley-III Fundamental Administration Video." Available at [<http://pearsonassess.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8027-23X&Mode=detail&Leaf=accessory&dsrc=015-8027-604#ISBN2>]. 2005.

Pearson Assessments. "Bayley-III Scoring Assistant® and PDA Administration Software." Available at [<http://pearsonassess.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8027-23X&Mode=detail&Leaf=software>]. 2005.

Tobin, Renee M., and Kathryn E. Hoff. "Review of the Bayley Scales of Infant and Toddler Development-Third Edition." In *The Seventeenth Mental Measurements Yearbook*, edited by Kurt F. Geisinger, Robert A. Spies, Janet F. Carlson, and Barbara S. Plake. Lincoln, NE: The Buros Institute of Mental Measurements, 2007.

Venn, John J. "Review of the Bayley Scales of Infant and Toddler Development-Third Edition." In *The Seventeenth Mental Measurements Yearbook*, edited by Kurt F. Geisinger, Robert A. Spies, Janet F. Carlson, and Barbara S. Plake. Lincoln, NE: The Buros Institute of Mental Measurements, 2007.

West, J., and C. Andreassen. "*Measuring Early Development in the Early Childhood Longitudinal Study-Birth Cohort.*" Paper presented at the Selecting Measures for Young Children in Large Scale Surveys, Washington, DC: Research Network on Child and Family Well-Being and the Science and Ecology of Early Development, 2002.

BEHAVIOR ASSESSMENT SYSTEM FOR CHILDREN, SECOND EDITION (BASC-2), 2004 (2011 Update)

Authors:

Cecil R. Reynolds and Randy W. Kamphaus

Publisher:

Pearson

(800) 328-2560

<http://www.pearsonschool.com/>

Instrument:

<http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAa30000>

Initial Material Cost:

BASC-2 Hand-Scored Starter Set (includes manual and 25 hand-scored forms of Teacher Rating Scale, Parent Rating Scale, Self-Report of Personality, Parent Feedback Report, Structured Developmental History, Student Observation System, Training Video): \$492.25

Starter set with English and Spanish materials: \$572.45

BASC-2 ASSIST Scoring & Reporting Software Starter Sets range from \$637.90 (English) to \$998.75 (English and Spanish)

Representativeness of Norming Sample¹:

The norming sample for the Teacher Rating Scales, Parent Rating Scales, and Self-Report of Personality included more than 13,000 2 to 18 year olds (400 were 2 to 3 years old, and 650 were 4 to 5 years old).

Assessments took place from 2002 to 2004. All age groups were half female and matched to targeted U.S. population estimates from the March 2001 Current Population Survey on socioeconomic status, race, and

¹ Authors also collected information from a clinical norming sample of 1,779 4 to 18 year olds whose parents had a clinical diagnosis or classification and from a sample of 19 to 21 year olds enrolled in special education classes (sample size not specified).

geographic region. Children with an emotional or behavioral disturbance or language impairment were slightly oversampled. The Spanish speaking sample is described below (see Other Languages).

Languages:

English, Spanish

Type of Assessment:

Parent/teacher report, observation, and self-report

Age Range and Administration Interval:

2 to 25 years old

Personnel, Training, Administration, and Scoring

Requirements: Personnel administering the BASC-2 vary with assessment type. Teachers administer the Teacher Rating Scale, which takes 10 to 15 minutes. Parents or guardians administer the Parent Rating Scale, which takes 10 to 20 minutes. A clinician administers the Student Observation System, which takes 15 minutes.

Training videos and DVDs are available on the publisher's web site for \$137.65.

Scoring times vary with the use of paper forms or computer software.

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 3 (all .65 or higher)

Validity: 3 (mostly .5 or higher for concurrent)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 2 (administered by teacher, clinician, and/or parent- or self-reported; scored by teacher or clinician only)

Description: The Behavior Assessment System for Children, Second Edition (BASC-2) evaluates the behavior of children and adults 2 to 25 years old. Information on children's behavior is triangulated from the perspective of self, teacher,

and parent by using five types of assessments or components: the Teacher Rating Scale, Parent Rating Scale, Student Observation System, Structured Developmental History, and Self-Report of Personality (for 6 to 25 year olds only).

Authors indicate that the components may be used individually or in combination.

The Teacher and Parent Rating scales (TRS and PRS) measure adaptive and problem behaviors in the community, school, and home setting. They consist of forms for three age groups: preschool (2 to 5 years old), child (6 to 11 years old), and adolescent (12 to 21 years old). For preschool-age children, the TRS includes 100 items and is administered by an adult caregiver or teacher who observes the child in a preschool setting and enjoys good rapport with the child. The PRS includes 134 items and is administered by the subject's parent or guardian by observing the child in a controlled setting (such as a clinician's office). Both the TRS and PRS are divided into Composite scales and Clinical and Adaptive scales. The four Composite scales for preschool-age children are Externalizing Problems, Internalizing Problems, and Adaptive Skills as well as the Behavioral Symptoms Index (BSI), which measures the broad level of problem behaviors. The Clinical scales include Hyperactivity, Aggression, Anxiety, Depression, Somatization, Attention Problems, Atypicality, and Withdrawal. The Adaptive scales include Adaptability, Social Skills, and Functional Communication. The PRS includes an additional Adaptive scale called Activities of Daily Living. The optional Content scales that help interpret the primary scales include Anger Control, Bullying, Development Social Disorders, Emotional Self-Control, Executive Functioning, Negative Emotionality and Resiliency. The Student Observation System (SOS) is a 14-item behavior

checklist administered by a clinician that primarily documents a child's classroom behaviors, including teacher-child interaction. The Portable Observation System (which incorporates the SOS form) is a computer- or PDA-based program that facilitates observations for the SOS. The Structured Developmental History (SDH) has 19 sections, such as the respondent's background information about family, medical history, relationships, interests, behavior, and educational history. A parent or guardian may complete the SDH in the home, school, or clinic setting. The developmental history/background obtained by the SDH informs the diagnostic or treatment process as fully as possible.

The BASC-2 updates the BASC (1992) with new content scales. The TRS and PRS now include Functional Communication, Activities of Daily Living, and Adaptability. The Self-Report of Personality includes Hyperactivity and Attention Problems. The BASC-2 also expands the age range and changes and some items. It includes a new college-level Self-Report of Personality, more detailed clinical norms, minor enhancements to the SDH, and technology-enhanced features such as the SOS for use with a laptop or PDA and optional scoring software.

Other Languages: The PRS, Self-Report of Personality, and SDH are available in Spanish. Item order, content, and administration are the same as in the English version. The Spanish-language version norming sample for the PRS included 82 Hispanic preschool-age children and

229 Hispanic children and adolescents. The sample was approximately half female and mostly from the North Central and Western regions of the country. The median internal consistency reliability alpha for the PRS scales was .71 among preschoolers (lower than the median alphas obtained for the English form and sample). For the PRS Composite scales, alphas ranged from .79 to .85 among preschool-age children. Intercorrelation coefficients were generally comparable between composites and scales of the PRS Spanish preschool forms and English preschool forms.

Uses of Information: The BASC-2 may be used to help diagnose emotional and behavioral disorders, identify special needs and educational program placement, conduct research on child emotional and behavioral disorders, evaluate interventions and programs, determine whether a defendant's behavior was attributable to a disability, and plead testimonies for child custody cases.

Reliability:

1. Internal consistency reliability: Cronbach's alphas for the TRS were reported for Composite and scale scores. Among 2 to 3 year olds, coefficients for the Composite scale coefficients ranged from .87 (Internalizing Problems) to .96 (BSI). The Clinical scale coefficients ranged from .75 (Anxiety) to .92 (Aggression and Attention Problems). The Adaptive scale coefficients ranged from .79 (Adaptability) to .87 (Social Skills). Among 4 to 5 year olds, coefficients for the Composite scales ranged from .91 (Internalizing Problems) to .96 (BSI). Clinical scale coefficients ranged from .81 (Anxiety) to .93 (Attention Problems). Adaptive scale coefficients ranged from .83 (Adaptability) to .88 (Social Skills).
2. Test-retest reliability: The TRS correlation coefficients for the preschool-age sample ranged from .83 to .92 for the Composite scales and from .72 to .92 for the primary scales. The tests were conducted with a sample of 69 preschoolers demographically similar to the norming sample, except that nearly half resided in the South. The same teacher scored the tests twice over an administration interval ranging from 8 to 65 days.

The PRS correlation coefficients for the preschool-age sample ranged from .79 to .86 for the Composite scales and from .66 to .88 for the Clinical and Adaptive scales. The same parent or guardian assessed 87 preschool-age children in the sample twice, with an interval of 9 to 70 days between ratings. The test-retest sample was demographically similar to the norming sample.

3. Inter-rater reliability: for the preschool-age TRS, the interval between teacher ratings ranged from 0 to 62 days. The adjusted median reliability estimate was .65 and ranged from .40 to .80 for Composite, Clinical, and Adaptive scale scores. The preschool-age sample size was 74, and the median age was 52 months. The group was demographically similar to the norming sample except that it was 68 percent male and more than half of the children resided in the Northeast. For the PRS, the interval between ratings ranged from 0 to 70 days. The adjusted median was .77 and ranged from .72 to .86 for Composite, Clinical, and Adaptive subscale scores. The preschool-age sample of 87 was similar to the norming sample except that nearly half of the children resided in the South.

Validity:

1. Content validity: The authors used Confirmatory Factor Analysis to examine the TRS. Factor loadings were high across age levels for the Externalizing Problems factor (Hyperactivity, Aggression, and Conduct Problems) and Adaptive Skills factor. The

Internalizing Problems factor loadings were in the middle to high range, with more variation across age groups. For the PRS, Covariance Structure Analysis factor loadings varied considerably across scales and age groups, although the Externalizing Problems composite had high factor loadings across age groups.

2. Concurrent validity: the TRS and the Achenbach System of Empirically Based Assessment (ASEBA) Caregiver-Teacher Report form for ages 1.5 to 5 years were completed for 46 children. Correlation coefficients between the TRS Composites and the ASEBA scale raw scores ranged from .64 (Internalizing Problems) to .82 (Externalizing Problems). Correlations between the BASC-2 Clinical and Adaptive scales and ASEBA scale raw scores ranged from .02 (BASC-2 Somatization with ASEBA Somatic Complaints) to .90 (BASC-2 Aggression with ASEBA Aggressive Behavior).

The PRS and the ASEBA Child Behavior Checklist were completed for a sample of 53 2 to 5 year olds. Correlation coefficients between the PRS Composites and the ASEBA scale raw scores ranged from .68 (Internalizing Problems) to .83 (Externalizing Problems). Correlations between the BASC-2 Clinical and Adaptive scales and the ASEBA scale raw scores ranged from .32 (BASC-2 Anxiety with ASEBA Anxious/Depressed) to .79 (BASC-2 Hyperactivity with ASEBA Attention Deficit Hyperactivity Disorder).

In addition, the authors compared the preschool forms of the BASC-2 TRS and PRS to the original BASC TRS and PRS. Most coefficients were .90 or higher, with the exception of Anxiety (.89), Atypicality (.83), and Adaptability (.86) in the TRS and Adaptive Skills (.85), Atypicality (.76), Attention Problems (.78), and Adaptability (.84) in the PRS.

3. Construct validity: correlations between the individual TRS and PRS scales for preschoolers measuring similar constructs were higher than those measuring distinct constructs.² Correlations between PRS and TRS Composite scales ranged from .17 (PRS Internalizing Problems with TRS Internalizing Problems) to .52 (PRS Adaptive with TRS Adaptive). Correlations between PRS Clinical scales and TRS Clinical scales ranged from .17 (PRS Depression with TRS Depression) to .43 (PRS Attention Problems with TRS Attention Problems). Correlations between PRS Adaptive scales and TRS Adaptive scales ranged from .28 (PRS Adaptability with TRS Adaptability) to .59 (PRS Functional Communication with TRS Functional Communication).

Bias Analysis: The authors removed fewer than five items from the TRS and PRS based on the results of partial correlations between

² Comparisons between scales with similar constructs are presented here. The manual also includes correlation coefficients between composites and scales of different constructs.

individual items and the demographic groups being compared and results of Differential Item Functioning analysis.

Methods of Scoring: The assessor hand scores paper versions of the PRS and TRS by summing the circled item response values, ranging from 0 and 3. The manual describes how to convert raw scores into T scores and percentile rankings by using the norm tables and then describes how to enter scores onto a Summary Table that helps identify high, average, and low scores compared with the mean T scores. The BASC-2 ASSIST and ASSIST Plus software may be used to score the Teacher and Parent Rating scales.

Interpretability: Assessors enter scores into a Summary Table that provides guidance on low or high scores and their meaning. Low or average scores are typical, and high scores are problematic.

Training Support: The publisher sells training videos and DVDs as either part of the BASC-2 Starter Set or separately.

Adaptations/Special Instructions for Individuals with Disabilities: Audio recordings of the items on the BASC-2 PRS are available for parents with reading problems.

Report Preparation Support: The BASC-2 ASSIST and ASSIST Plus software programs provide reports of the Teacher and Parent Rating scales. The programs generate profiles, calculate validity indices, identify strengths and weaknesses, and compute multirater comparisons. ASSIST Plus also reports on the optional Content scales and target behaviors for intervention and helps

with diagnosis of possible disorders such as those in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV-TR). Clinicians record Teacher and Parent Rating scale scores on Parent Feedback Reports and provide parents with an overview of the test process and scale content as well as with an interpretation of the child's scores.

References:

Pearson. *BASC-2 ASSIST Scoring & Reporting Software*. 2004.

Pearson. *BASC-2 Training Videos*. 2004.

Reynolds, Cecil R., and Randy W. Kamphaus. *Behavior Assessment System for Children, Second Edition Manual*. Circle Pines, MN: AGS Publishing, 2004.

Stein, Stephanie. "Review of the Behavior Assessment System for Children, Second Edition." In *The Seventeenth Mental Measurements Yearbook*, edited by Kurt F. Geisinger, Robert A. Spies, Janet F. Carlson, and Barbara S. Plake. Lincoln, NE: The Buros Institute of Mental Measurements, 2007.

Watson, T.S., and Katherine Wickstrom. "Review of the Behavior Assessment System for Children, Second Edition" In *The Seventeenth Mental Measurements Yearbook*, edited by Kurt F. Geisinger, Robert A. Spies, Janet F. Carlson, and Barbara S. Plake. Lincoln, NE: The Buros Institute of Mental Measurements, 2007.

BRIGANCE INVENTORY OF EARLY DEVELOPMENT II STANDARDIZED (IED-II STANDARDIZED), 2010¹ (2011 Update)

Authors:

Albert H. Brigance and Frances Page Glascoe

Publisher:

Curriculum Associates

(800) 225-0248

<http://www.curriculumassociates.com/default.asp>

Instrument:

<http://www.curriculumassociates.com/products/detail.asp?title=BrigIED2-CA>

Initial Material Cost: IED-II Standardized Kit (standardized assessments, Standardization and Validation Manual, 20 Record Books, testing accessories kit, canvas tote): \$299

No outside scoring is required. However, an IED-II Standardized scoring tool is available online for free. The Online Management System costs \$129 or \$159 for a 1-year license.

Representativeness of Norming Sample: The IED-II Standardized was normed in 2003, using a sample of 1,171 children age 0 to 5-years-old and older from public and private schools, day care centers, and preschools. Two-thirds of the sample was 3-years-old or younger. Based on the 2003 U.S. Census, the norming sample was stratified on ethnicity, parental education attainment, income level, location of residence (urban/suburban versus rural), and marital status of parent. The sample was approximately half male and included all four major geographic regions of the United States, with slightly lower representation from the North.

¹ The IED-II Standardized includes 44 assessments from the Inventory of Early Development II (IED II) that have been standardized.

Languages:

English

Type of Assessment:

Observation, parent interview, parent report, teacher report, and direct child assessment

Age Range and Administration Interval:

Children age 0 to 7 years old

Personnel, Training, Administration, and Scoring Requirements: Administered by a teacher, school psychologist or developmental expert, or other early education professionals.

The manual suggests that the assessor familiarize him- or herself with the assessments and practice administration a few times before assessing a child. Specific directions accompany each assessment and should be closely followed. The publisher's web site offers free e-training on how to administer and score the instrument and consists of 5 short video modules.

If all the assessments are administered, the total time for administration is approximately 20 minutes for infants and 50 minutes for most 4- and 5-year-olds. Scoring time varies with the number of assessments administered.

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 3 (all .65 or higher)

Validity: 3 (mostly .5 or higher for concurrent)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 2 (administered and scored by a child development professional familiar with the instrument)

Description: The Brigance Inventory of Early Development II Standardized (IED-II Standardized) measures children's strengths and weaknesses across basic skill areas such as

physical development, language development, early numeracy and emergent literacy skills, social-emotional development, and self-care daily living skills. The IED-II Standardized is

designed for use with children from birth to 7 years old and may be administered by teachers, school psychologists, developmental experts, and other early education professionals. Each assessment includes directions specifying whether the assessment method is a parent interview, child observation, and/or performance (i.e., behaviors/actions elicited from children). Five domains cover 10 subdomains, 8 composites, 44 assessments, and 706 items. Assessments may be administered in any order, and assessors may select assessments from any domain for administration. The manual provides guidelines for determining which assessments to administer based on the child's age and needs. Most are appropriate for preschool and kindergarten-age children, with four assessments used exclusively for infants and toddlers. Some assessments require manipulatives such as toys, objects to count, and colored blocks. These and other manipulatives such as pencils, scissors, and crayons are included in the Accessories Kit available from the publisher.

The five domains are Physical Development, Language Development, Academic/Cognitive, Daily Living, and Social and Emotional Development. Physical Development subdomains include Fine-Motor Skills and Gross-Motor Skills (13 assessments and 149 items). The Fine-Motor Skills subdomain includes a Drawing/Visual Motor Composite and a Writing Skills Composite. The Gross-Motor Skills subdomain includes a Nonlocomotor Composite and a Locomotor

Composite. Language Development subdomains are Receptive Language and Expressive Language (11 assessments and 180 items). Within the Receptive Language subdomain, the two composites are (1) Nouns and Early Listening and (2) Actions. Within the Expressive Language subdomain, the two composites are Isolated Skills and Contextual Skills. In the Academic/Cognitive domain, the two subdomains are Mathematical/General Concepts and Literacy (13 assessments and 236 items). In the Daily Living domain, the two subdomains are Self-Help and Prevocational (5 assessments and 64 items). In the Social and Emotional Development domain, the two subdomains are Play Skills and Behaviors and Engagement and Initiative Skills (2 assessments and 42 items). In the Self-Help subdomain and the Social-Emotional domain, there are 35 items each on the Parent's Report Form and the Teacher's Report and Scoring Form (one or the other is used, not both).

Compared to the IED's 2008 version, the current version offers more resources for users, including objectives for writing Individualized Education Programs (IEP) at the end of every assessment and new utilities such as online versions of the Self-Help subdomain and Social and Emotional Development domain (paper versions are available in the manual as well), the IED-II Standardized Scoring Tool, and a Chronological Age Calculator.

Other Languages: Items were field tested in Spanish, but no IED-II Standardized Spanish

language materials are available. Brigance Screens II, shorter assessments that cover some of the same skills as the IED-II Standardized, are available in Spanish.

Uses of Information: Results may be used to identify and diagnose developmental delays or to make recommendations for additional testing. They also inform instruction and special education referral decisions as goals in an IEP.

Reliability:

1. Internal consistency reliability: Guttman Lambda coefficients for each composite, subdomain, and domain all exceeded .85. The total Adaptive Behavior coefficient, which is the sum of all five domains, was .99.
2. Test-retest reliability: estimates are available from the 1991 standardization sample and 2003 norming sample. The latter estimates are based on two administrations within one week for 36 children in the birth to 12 month old group because new test items applied to this age group. Combined results from the two studies reveal coefficients ranging from .68 to .99 across age groups and domains. The composite Total Adaptive Behavior scores by age group ranged from .89 (61 months and older) to .95 (0 to 12 months).
3. Inter-rater reliability: in 1988, two teachers assessed 20 5- and 6-year-olds enrolled in an early intervention program. The 2003 study data augmented the earlier data, with combined results from both studies showing

percentages of agreement ranging from 80 to 96 percent across the five domains.

Validity:

1. Content validity for the IED-II Standardized is based on an extensive literature review by the author, collaboration with child development professionals from other states who helped with item selection, and field testing in 16 states. A factor analysis using varimax rotation conducted between the subdomains resulted in three factors: (1) Understanding and Expressing, (2) Movement and Social Activity, and (3) Academic/Preacademic. The analysis revealed the highest factor loadings between Contextual Skills and Understanding and Expressing (.88) and the lowest factor loadings between Drawing/Visual Motor and Movement and Social Activity (-.30).
2. Concurrent validity: some 484 children from the standardization sample took several batteries of additional assessments to test for concurrent validity with each of the IED-II Standardized subdomains. Results were aggregated across age groups and reported as correlations between IED-II Standardized subdomains or domains and batteries of diagnostic measures.

The Cognitive Skills battery consisted of the Bayley Scales of Infant Development and the cognitive portions of the Battelle Developmental Inventory. Children 2 years old and older also took the Slosson

Intelligence Test Revised (SIT-R). The aggregated cognitive skills battery (all ages) correlated at .88 with the IED-II Standardized Cognitive domain.

For infants and toddlers, the Language Development and Adaptive Behavior battery included the Rosetti Infant Toddler Language Scale; Preschool Language Scale; Receptive-Expressive Emergent Language Test, second edition (REEL-2); Sequenced Inventory of Communication Development; Battelle Developmental Inventory Screening Test; or Communication Domain of the Vineland Adaptive Behavior Scale. Parents of children 2 years old and older completed the Child Development Inventory (CDI). The receptive language battery (all ages) correlated at .64 with the IED-II Standardized Receptive Language subdomain. The expressive language battery (all ages) correlated at .54 with the IED-II Standardized Expressive Language subdomain.

For Physical Development, infants and toddlers took either the motor domain of the Vineland or the motor domain of the Bayley Scales, motor portions of the Battelle, or the Alberta Infant Motor Scale. Children 2 years old and older took the CDI. The fine motor diagnostic battery (all ages) correlated at .74 with the IED-II Standardized Fine Motor subdomain. The gross motor battery (all ages) correlated at .70 with the IED-II Standardized Gross Motor subdomain.

For Social and Emotional Development, infants and toddlers took the Bayley Infant Behavior Record and Interaction Scale from the Rosetti and/or the Social domain from Vineland. The social and emotional battery (all ages) correlated at .51 with the IED-II Standardized Social and Emotional Development domain.

The IED-II Standardized distinguishes between children with the following characteristics based on Chi-square significance testing between children with and without a characteristic: (1) premature birth, (2) psychosocial risk, (3) developmental disabilities, and (4) highly advanced development. Descriptive data show that standardized assessment scores increase with age, although no significance testing was conducted. Analysis of covariance analyses revealed performance differences between children with parents who had and had not completed high school. In addition, children who qualified for free and reduced-price lunch and/or Medicaid did not perform as well as children from higher-income families.

3. Predictive validity: in the absence of studies on the IED-II Standardized, the author cites the measure's predictive validity based on "high correlations" between Brigance Screens and measures of academic achievement, intelligence, language development, family environment, and teacher ratings six months to two years later.

4. Construct validity: the author evaluated intercorrelations between IED-II Standardized domains, composites, and broader skill areas that measured similar and different capabilities. Correlations ranged from .36 (Social and Emotional Development correlated with Writing) to 1.0 (Total Physical Development correlated with Total Motor).

Bias Analysis: No information available.

Methods of Scoring: For each assessment, the assessor sums the total number of correct items in each assessment, including all items below the basal, to obtain a raw score. Raw scores are totaled by composite, subdomain, and domain on the Standardized Scoring Sheet. Appendix tables in the manual convert raw composite, subdomain, and domain scores into quotients, percentiles and age equivalents, and instructional ranges.² If four or five domains are completed, a Total Adaptive Behavior score may be calculated by finding the median domain quotient among all quotients for all domains that have been computed. The Online Management System provides a Total Adaptive Score and standardized scores for domains, subdomains, and composites.

Interpretability: Developmental profiles in individual children's Record Books may be used to plot quotients for composite, subdomain, and domain scores. The assessor interprets the

plotted quotients by referring to a graph showing average ranges on the Standardized Scoring Sheet. Plotted quotients are within or outside the ranges. The results may be conveyed to parents and caregivers. Progress toward children's mastery of assessed skills and objectives for meeting un-mastered skills may be tracked in Record Books or as part of a child's IEP. The end of every assessment provides suggestions on how to phrase IEP goals with practitioners and parents.

Training Support: The publisher offers free online training modules on how to administer and score the instrument. The modules comprise a five-video series called CA 101 that provides step-by-step instructions.

Adaptations/Special Instructions for Individuals with Disabilities: The manual includes suggested adaptations for children with motor impairment, hearing impairment, vision impairment, speech impairment, emotional disturbance, significant health problems, autism and developmental disorders, traumatic brain injury, and intellectual disabilities.

Report Preparation Support: The Online Management System provides downloadable IEP goals, teacher and administrator reports and communications for families, and tools to meet reporting requirements for the Individuals with Disabilities Education Act (IDEA) and state-specific content standards.

² Instructional ranges are derived from age equivalents in months for the raw score ranges.

References:

Brigance, Albert H. *Brigance Diagnostic Inventory of Early Development II*. North Billerica, MA: Curriculum Associates, 2008.

Brigance, Albert H. *Brigance: Inventory of Early Development II, Standardized*. North Billerica, MA: Curriculum Associates, 2010.

Glascoe, Frances P. *Brigance: Inventory of Early Development II, Standardization and Validation Manual*. North Billerica, MA: Curriculum Associates, 2010.

CAREY TEMPERAMENT SCALES (CTS), 2000

Authors:

William B. Carey

Publisher:

Behavioral-Developmental Initiatives
(800) 405-2313
www.b-di.com

Initial Material Cost:

Specimen set: \$60 (includes a sample of each of the five CTS questionnaires, with scoring and profile sheets, and the Test Manual)

Representativeness of Norming Sample:

Not nationally representative; it was normed on primarily a white middle class population living in the eastern United States.

Languages:

English

Type of Assessment:

Parent report

Age Range and Administration Interval:

1 month to 12 years

Questionnaires are available for the following age ranges: 1 to 4 months, 4 to 11 months, 1 to 2 years, 3 to 7 years, 8 to 12 years

Personnel, Training, Administration, and Scoring Requirements:

Intended for professional use by persons licensed or certified to provide care to children and their parents; administration time is 20 minutes and requires an early high school reading level; scoring time is 15 to 20 minutes for hand scoring and 4 minutes for computer scoring; a professional is needed for scoring and interpretation.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 2 (all or mostly under .65)

Validity: 1 (none described)

Norming Sample Characteristics: 2 (not nationally representative, normed within the past 15 years)

Ease of Administration and Scoring: 2 (self-administered, but scored by a professional)

Description: The Carey Temperament Scales (CTS) are sets of items for obtaining parent report of a child's temperament. It can be used in research and/or clinical practice. The CTS uses 75 to 100 descriptions of behavior to assess the 9 New York Longitudinal Study characteristics of temperament: (1) activity level, (2) rhythmicity, (3) approach-withdrawal, (4) adaptability, (5) intensity, (6) mood, (7) attention span and persistence, (8) distractibility, and (9) sensory threshold. CTS is comprised of 5 different questionnaires, three of which are particularly relevant to Early Head Start programs. These are the Early Infant Temperament Questionnaire

(EITQ) for infants ages 1 to 4 months, the Revised Infant Temperament Questionnaire (RITQ) for infants ages 4 to 8 months (and applicable, but not normed, for ages 9 to 11 months), and the Toddler Temperament Scale (TTS) for children ages 1 to 2 years. The other two scales, the Behavioral Style Questionnaire (BSQ) and Middle Childhood Temperament Questionnaire (MCTQ) are for children ages 3 to 12. Each questionnaire contains up to 100 items that are rated on a 6-point scale of frequency ranging from almost never to almost always. These instruments are designed for caregivers who spend a substantial amount of time with the child being assessed.

Uses of Information: The CTS can help caregivers understand a child's temperament and behavioral style. The scales can also be used by caregivers to help place a child in an environment more suitable to the child's temperament or to adapt the environment (including the home and parenting strategies) to the child's temperament. Temperament itself is not considered amenable to intervention.

Reliability: (1) Internal consistency (Cronbach's alphas): EITQ: scale ranged from .43 to .76 (median = .62); RITQ: scale ranged from .49 to .71 (median = .57); TTS: scale ranged from .53 to .86 (median = .70); BSQ scale ranged from .47 to .80 (median = .70); MCTQ scale ranged from .71 to .83 (median = .82). (2) Test-retest reliability: EITQ (20 day test interval): scale ranged from .64 to .79 (median = .68); RITQ (25 day interval): scale ranged from .66 to .81 (median = .75); TTS (1 month interval): scale ranged from .69 to .89 (median = .81); BSQ (1 month interval): scale ranged from .67 to .94 (scale median = .81); MCTQ (75 day interval): scale ranged from .79 to .93 (median = .88).

Validity: Literature on the clinical evidence for validity and appropriate use of temperament data in practice can be found in *Coping with Children's Temperament* (1995), written by Carey and McDevitt or in *Developmental-Behavioral Pediatrics* (1992), edited by Levine, Carey, and Crocker.

Method of Scoring: The CTS can be hand or computer scored. Items are tabulated to yield a category score for each of the nine areas, which

are then compared to the norms for the category. If using the software, the Professional Report includes the temperament profile, raw and standardized scores, individualized interpretive report and validity checks for social desirability, missing data and ratings/perceptions discrepancies. The Caregiver Report contains the temperament profile and an interpretive report of scores written for the caregiver and personalized with the child's name and gender.

Interpretability: Category scores for each of the nine areas can be compared to norms for the category. The manual gives instructions for interpreting the results, depending on whether or not the computer or manual scoring is used. In addition, the authors stress the importance of supplementing the results from the CTS with information gathered from interviews, observations, and other information collected by trained professionals.

Training Support: CTS practice sets are available through the publisher. Individuals with questions may email publisher.

Adaptations/Special Instructions for Individuals with Disabilities: None described. However, a younger age questionnaire can be used for individuals with mild delays.

Report Preparation Support: The manual states that the written report or profile should not be automatically given to caregivers because they may not have sufficient information about the limitations of the information. The professional should exercise his or her judgment when

deciding whether or not to share the written report or

References:

Carey, William B. *The Carey Temperament Scales Test Manual*. Scottsdale, AZ: Behavioral-Developmental Initiatives, 2000.

Carey, W.B. and McDevitt, S.C. *Coping With Children's Temperament: A Guide for Professionals*. New York: Basic Books, 1995.

Levine, M.D., W.B. Carey, and A.C. Crocker. *Developmental-Behavioral Pediatrics*, 2nd ed. Philadelphia, PA: W.B. Saunders, 1992.

THE CAROLINA CURRICULUM FOR INFANTS AND TODDLERS WITH SPECIAL NEEDS (CCITSN), ASSESSMENT LOG, SECOND EDITION, 1991

Authors:

Nancy Johnson-Martin, Kenneth Jens, Susan Attermeier, and Bonnie Hacker.

Publisher:

Paul H. Brookes Publishing Co.
www.brookespublishing.com
1-800-638-3775

Initial Material Cost:

\$43

Representativeness of Norming Sample:

None (criterion-referenced)

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

0 to 2 developmental years with mild to severe special needs

Personnel, Training, Administration, and Scoring Requirements:

Can be used by both professionals and paraprofessionals. Effort made to avoid use of technical jargon in materials to encourage broader usage. Requires informal observation period in which the examiner can assess and score child in about 15 to 20 minutes. The Assessment Log is designed to be used in conjunction with the Carolina Curriculum.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described)

Validity: 1 (content validity only described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a professional or paraprofessional)

Description: The Carolina Curriculum for Infant and Toddlers with Special Needs (CCIT-SN) is designed for use with infants from birth to 2 years developmental age who have mild to severe special needs. The curriculum covers 6 developmental domains (cognition, communication, social/adaptation, fine motor, and gross motor) that are divided into 26 teaching areas (or sequences). CCITSN has an Assessment Log that enumerates 26 sequences and the specific skills under each sequence are ordered sequentially in terms of the expected development of children. The number of items a child is assessed on is at the discretion of a professional or paraprofessional, who administers the CCITSN Assessment Log through an informal observation of the parent-child interaction. This format is

preferred to a more clinical, structured approach to assessment.

Uses of Information: The CCITSN Assessment Log is used to identify the curriculum entry point, to inform the intervention plan, and to monitor progress in accomplishing the skills covered by the curriculum.

Reliability: None described.

Validity: Content validity: the selection of items for inclusion in the curriculum Assessment Log was accomplished through a review of norm-referenced tests of development. A multi-disciplinary panel of specialists helped in the final selection process. The curriculum was field tested in 22 intervention programs in North Carolina

and in 10 national sites, and the interventionists found it to be useful both for assessing infants with disabilities and for developing their intervention programs.

Method of Scoring: Behaviors on the question items are scored as either typical of the child (+), emergent (+-), or never observed (-). Once behaviors have been scored, the Developmental Progress Charts are used to chart assessment results and develop a profile of the child's skills. There is a blank box on the Developmental Progress Chart to correspond to each item on the Assessment Log, which is colored in completely when an item is passed. If the skill is inconsistently performed or emerging, the space is partially colored in.

Interpretability: Generally, a child's performance can be assessed based on his or her performance of the three items after the first failure and the three items before the first success in each sequence. The Assessment Log then serves as a basis for intervention using the Curriculum Sequences by selecting the first activity the child failed or that was just emerging in each sequence.

Training Support: "Brookes on Location" professional development seminar, Using The

Carolina Curriculum to Assess and Intervene with Young Children with Special Needs, is available through the publisher.

Adaptations/Special Instructions for Individuals with Disabilities: The CCITSN is designed specifically to optimize the development of children with mild to severe special needs. Interventions, and their associated assessment, are tailored to the child's impairment, and the standard approach will be modified if a handicapping condition makes it inappropriate. The curriculum has special needs options and adaptations available for those with vision, motor or hearing needs.

Report Preparation Support: A profile of the child's skills can be obtained by completing the Developmental Progress Chart.

References:

Johnson-Martin, Nancy, Kenneth Jens, Susan Attermeier, and Bonnie Hacker. 2001. The Carolina Curriculum for Infants and Toddlers with Special Needs, Second Edition. Baltimore, MD: Paul H. Brookes Publishing Co.

DENVER II DEVELOPMENTAL SCREENING TEST (DDST-II), 1989

Authors:

William K. Frankenburg and J.B. Dodds

Publisher:

Denver Developmental Materials, Inc.
1-800-419-4729 • 303-355-4729
www.denverii.com

Initial Material Cost:

Denver II Test Kit (includes 100 forms, Screening Manual, and test items): \$84 Training Video: \$215 (purchase) or \$90/week (on-site rental)

Representativeness of Norming Sample:

The English version of the test was normed from 1987 to 1989 on a quota sample of 2,096 English-speaking children in Colorado with no obvious special needs. These children were of varying ages (between 2 weeks and 6.5 years), levels of maternal education, places of residence and cultural backgrounds. The Denver norming sample is representative of Colorado children (from 1980 US Census), and slightly overrepresents Hispanic infants, and underrepresents African American infants. However, when comparing the Colorado average 90% norms with the theoretical US composite norms, there were no clinically significant differences. The DDST-II Spanish version was not normed on Spanish-speaking children, but is a direct translation.

Languages:

English and Spanish (translation of directions and test forms)

Type of Assessment:

Direct child assessment and parent report

Age Range and Administration Interval:

Birth to 6 years

Personnel, Training, Administration, and Scoring Requirements:

Administration of the DDST-II and scoring of the DDST-II and PDQ-II require a professional or paraprofessional. The manual suggests that users carefully review the manual, review the training videotape, and practice testing children of various age groups in order to properly administer and interpret the DDST-II. A two-day training is also suggested. The test takes 10 to 20 minutes to administer, and 1 to 2 minutes to score. The Prescreening Developmental Questionnaire (PDQ-II) takes about 10 to 15 minutes to complete.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (both inter-rater and test-retest reliability have high percent agreement)

Validity: 1 (none described)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (scoring the DDST-II and the PDQ-II requires a highly trained individual)

Description: The DDST-II is a 125-item standardized measure that is designed to determine whether a child's development is within the normal range. It includes a set of questions for parents and tests for the child on twenty simple tasks and items that fall into four sectors: Personal-Social (25 items), Fine Motor Adaptive (29 items), Language (39 items), and Gross Motor (32 items). The number of items administered

during an assessment will vary with the child's age and ability.

A Prescreening Developmental Questionnaire (PDQ-II) has been developed to help parents quickly identify whether their children need further assessment. The PDQ-II is a pre-screening consisting of 91 parent questions from the DDST-II. It was created for parents to complete easily and quickly to assess whether their children have

non-normal scores and need to complete the full DDST-II. The PDQ-II was revised in 1998 and uses the norms developed for the DDST-II. The questionnaires are divided by age range (0 to 9 months, 9 to 24 months, 2 to 4 years, and 4 to 6 years).

Uses of Information: The DDST-II is intended for use as a screening tool to detect developmental delays. The DDST-II provides a clinical impression of a child's overall development and confirms suspected potential developmental difficulties with an objective measure. It can be used to determine how a child compares to other children and identify children for whom additional in-depth assessment should be conducted. The authors do not recommend using it to predict later development status, as an in-depth assessment of developmental functioning, or to plan individual intervention programs.

Reliability: (1) Internal consistency reliability: no information available. (2) Test-retest reliability: 89 percent agreement between test scores for a 7- to 10-day interval between test administrations by the same tester. (3) Inter-rater reliability: for the standardization sample, percentage agreement between examiners and a criterion observer (inter-rater reliability) ranged from 92 to 98 percent.

Validity: No information available

Method of Scoring: The child's responses are recorded as Pass or Fail on the score sheets. The responses are examined to see if they fall into or

outside the normal expected range of success on that item for the child's age (the child is either classified as normal range, suspect, or delayed).

Interpretability: The DDST-II scoring process, which is described in the screening manual, requires that the individual test items be interpreted before the entire test is interpreted. The individual items are classified as: Advanced, Normal, Caution, Delayed, and No Opportunity. The category descriptors for the entire test include: Normal, Abnormal, Questionable, and Untestable.

Training Support: It is suggested that screeners be properly trained and pass the proficiency test before using the DDST-II for clinical purposes. There is a two-day training workshop offered (and outlined in the technical manual).

Adaptations/Special Instructions for Individuals with Disabilities: None mentioned

Report Preparation Support: None mentioned

References:

Frankenburg, William K. and J.B. Dobbs. Denver Developmental Screening Test II-Screening Manual. Denver: Denver Developmental Materials, 1990.

Frankenburg, William K. and J.B. Dobbs. Denver Developmental Screening Test II-Technical Manual. Denver: Denver Developmental Materials, 1990.

Frankenburg, William K. and J.B. Dobbs.
Denver Developmental Screening Test II-
Training Videotape. Denver: Denver
Developmental Materials, 1993.

Personal correspondence with Beverly
Bresnick, DDST-II technical expert and trainer,
July 10, 2002.

DEVELOPMENTAL OBSERVATION CHECKLIST SYSTEM (DOCS), 1994

Authors:

Wayne Hresko, Shirley Miguel, Rita Sherbenou, and Steve Burton

Publisher:

Pro-ed
(800) 897-3202
<http://www.proedinc.com/customer/productView.aspx?ID=826>

Initial Material Cost:

Complete DOCS Kit: \$129 (includes Examiner's Manual, 25 Cumulative Profile/Record Forms, 25 Developmental Checklist Profile/Record Forms, 25 Adjustment Behavior Checklist Profile/Record Forms, and 25 Parent Stress and Support Checklist Profile/Record Forms)

Representativeness of Norming Sample:

DOCS was normed on more than 1,400 children ages birth through 6 years from more than 30 states. Although a random sampling procedure was not used, characteristics of the normative group approximate those for the 1990 U.S. Census data relative to gender, geographic region, race/ethnicity, and urban/rural residence. The tests were conducted between November 1989 and December 1992.

Languages:

English

Type of Assessment:

Parent or caregiver report

Age Range and Administration Interval:

Birth to 6 years

Personnel, Training, Administration, and Scoring Requirements:

Examiners should have some training in administering and interpreting assessment instruments. The instrument can be completed by a parent with a fourth grade reading level. It takes 30 minutes to complete and 15 to 20 minutes to score all three checklists.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 2 (about half of the coefficients were < .5; about half were > .5)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (self-administered and scored by a trained individual)

Description: The Developmental Observation Checklist System (DOCS) is a three-part instrument to assess the development of very young children, their ability to adjust to their environment, and the level of stress and support in their environment. Part I uses the Developmental Checklist (DC), a parent report questionnaire, to assess the child's general development in the areas of cognition, language, social, and motor domains. It is answered in a yes/no format. Part II uses the Adjustment

Behavior Checklist (ABC) to screen for any problematic behaviors in the child's ability to adapt to his/her environment. Part III uses the Parental Stress and Support Checklist (PSSC) to identify family stress regarding the child and support used to mediate the stressors. Both the ABC and the PSSC are scored on a 4-point Likert-type scale.

Uses of Information: The DOCS is used to (1) identify infants and children with developmental delays or deficits in cognitive,

language, social, and motor abilities; (2) assess adjustment behavior; (3) determine levels of familial stress and support; (4) facilitate the proper professional referral for the child; (5) serve as a measurement device in research studies, (6) give direction to instructional practice, and (7) document educational progress.

Reliability: (1) Internal consistency reliability (Cronbach alphas): the alphas for age groups between birth and 3 years old were in the mid to high .90s for the DC components, in the .80s for the ABC, and in the low to mid .90s for the PSSC. (2) Test-retest reliability (with a 14- to 21-day interval): coefficients for children ages 2 to 3 ranged from .85 to .91 for the DC component and overall checklists and were .94 and .89 for the ABC and PSSC, respectively. (3) Inter-rater reliability: Parent to caregiver standard score reliability coefficients on the DC component and overall ranged from .91 to .94.

Validity: (1) Concurrent validity: The DC component quotient scores correlation with the Bayley Scale for Infant Development, Expressive One-Word Picture Vocabulary Test (EOWPVT), Denver Developmental Screening Test-Revised, McCarthy, Receptive-Expressive Emergent Language Test (REEL), Test of Early Language Development-2 (TELD-2), Stanford Binet (SB-4th), Slosson Intelligence Test-Revised (SIT-R), Test of Early Socioemotional Development (TOESD), and Vineland Adaptive Behavior Scale ranged from .35 (Developmental Language Quotient, Developmental Cognition Quotient, and Developmental Cognition Quotient with the

McCarthy) to .83 (Developmental Language Quotient with the TELD-2). The correlation of DC component quotient scores with the Parental Stress Inventory (PSI) ranged from -.72 to -.38. The correlations of the ABC and PSSC with the TOESD were .65 and .47, respectively. For the Vineland, the correlations with the ABC and the PSSC were .69 and .51, respectively. The correlations of the ABC and PSSC with the PSI were -.38 and -.72, respectively. The DOCS was also able to differentiate between children with normal development and those with developmental challenges. These validity tests were performed on children between the ages of 3 and 6. (2) Predictive validity: no information available.

Method of Scoring: To score the DC, the examiner needs to find the child's basal and ceiling points. The basal is established when the parent or caregiver marks "Yes" for five items in a row, and the ceiling is established when the parent or caregiver marks "No" for five items in a row. The DC score is the sum of all of the items below the basal (including the five basal items) and the number of "Yes" responses above the basal and below the ceiling. The ABC and the PSSC have no basals or ceilings. Checkmarks in each column are weighted according to their placement in the scale (responses are assigned a number between 1 and 4) and multiplied by a factor indicated on the response sheet. To compute a raw score for each checklist, the correct responses are summed. Using tables in the manual, the DC component checklist raw scores can be converted into

percentiles, standard scores, quotients (a distribution with a mean of 100 and a standard deviation of 15), normal curve equivalents, and age-equivalents (the child's performance age). The manual also has tables to convert the ABC and PSSC raw scores into percentiles and quotients.

Interpretability: The manual provides guidelines for interpreting DOCS scores, as well as cautions about their limitations. In general, while low DOCS scores may indicate the presence of developmental or environmental issues, they do not provide information on the sources and nature of the issues. The examiner is advised to always consider other sources of information, but especially when the assessment has practical implications for the child.

Training Support: None

Adaptations/Special Instructions for Individuals with Disabilities: Instructions are given for how to administer the instrument if the

individual is blind, illiterate, or below a 4th-grade reading level. In addition, the norms provided are appropriate for normally developing children such as those used in the standardization sample. If an individual's performance is to be compared with a more specific reference group (for example, deaf, retarded, or children older than 6), the authors state that the suitability of the DOCS for that group should be established before evaluating test performance.

Report Preparation Support: There are instructions in the manual for how to share the results with others, including parents.

References:

Hresko, W.P., S.A. Miguel, R.J. Sherbenou, and S.D. Burton. Developmental Observation Checklist System: A Systems Approach to Assessing Very Young Children Examiner's Manual. Pro-Ed: Austin, TX, 1994.

DEVELOPMENTAL PROFILE 3 (DP-3), 2007 (2011 Update)

Authors:

Gerald D. Alpern, Ph.D.

Publisher:

Western Psychological Services

(800) 648-8857

<http://portal.wpspublish.com>

Instrument:

http://portal.wpspublish.com/portal/page?_pageid=53,186601&_dad=portal&_schema=PORTAL

Initial Material Cost:

Kit with unlimited-use scoring and interpretation CD (includes 25 Interview Forms, 25 Parent/Caregiver Checklists, manual, CD): \$375 (\$210 without CD)

Spanish Interview Form (package of 10): \$30

Spanish Parent/Caregiver Checklist (package of 10): \$30

Representativeness of Norming Sample:

The sample consisted of 2,216 children from the four major Census Bureau regions. Fifty-nine interviewers were recruited from 21 states across the four regions. Children's ages ranged from 0 to 12 years, 11 months, with approximately 67 percent age 5 years, 11 months or younger. The sample was half female and demographically similar to the U.S. population, based on 2005 Census data.

Languages:

English, Spanish

Type of Assessment:

Parent interview or self-report

Age Range and Administration Interval:

From 0 to 12 years of age

Personnel, Training, Administration, and Scoring

Requirements: The Interview Form is administered by someone familiar with psychological or educational testing or by a paraprofessional. The Parent/Caregiver Checklist should be completed by a caregiver who knows the child well and can read and comprehend at the grade 6 level. A clinician or a professional with training in child development, psychology, and/or education should undertake scoring and interpretation.

Those administering the Interview Form should be familiar with the manual. The interview takes 20 to 40 minutes.

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 3 (mostly .65 or higher)

Validity: 3 (mostly .5 or higher for concurrent; all .4 or higher for predictive)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (self-administered or administered by someone with basic clerical skills; scored by clinician or designated assessor)

Description: The Developmental Profile 3 (DP-3) is both a screening tool and a diagnostic instrument designed to measure child development and functioning of children 0 to 12 years of age. The DP-3 may be administered as either an interview or a parent checklist. The content in both formats is identical; however, the

Interview Form question wording is slightly different, and the author favors the interview format because assessors may clarify questions throughout administration. The forms comprise five scales: Physical (35 items), Adaptive Behavior (37 items), Social-Emotional (36 items), Cognitive (38 items), and Communication (34 items). A

composite score called General Development (GD) is based on the five scales, although single scales may be administered because each scale was normed separately. On the Interview Form, each scale is adaptive such that items are arranged in order of developmental difficulty. The assessor must first calculate the child's chronological age and then find the correct item to start the interview based on the child's age. For children age 2 years and older, the assessor must obtain a basal level of five "yes" responses. If the assessor does not obtain the basal level, he or she tests backward from the start item until achieving five consecutive "yes" responses. The ceiling applies when five consecutive items are scored "no." For the Parent/Caregiver Checklist, the respondent answers all questions. The DP-3 updates the DP-II (1980) with norm-based standard scores, a Parent/Caregiver Checklist, an expanded age range, updated item content and interpretation guidelines, suggestions of intervention activities, a scoring and interpretation software program, and an overall GD score.

Other Languages: The Parent/Caregiver Checklist and Interview Form are available in Spanish. Information on the norming sample, reliability, validity, and English language equivalence and comparability is unavailable.

Uses of Information: The DP-3 may be used as a screening tool for health care providers. It may also be used to measure a child's progress in outcomes over the year and for research purposes. The author suggests use of the DP-3 as a diagnostic instrument to determine eligibility for

special education services or to plan an Individualized Education Program if supplemented with other information.

Reliability:

1. Internal consistency reliability: adjusted split-half reliability estimates were reported for the Interview Form by using the standardization sample by year of age and subscale. For ages 0 and 1 year, Pearson's correlations for subscales ranged from .84 to .93, and the composite GD score coefficient was .97. For ages 2 and 3 years, coefficients ranged from .82 to .88 for subscales, and the GD score coefficient was .95. For ages 4 and 5 years, coefficients ranged from .71 to .86 for subscales, and the GD score coefficient was .92. The report of internal consistency for the Parent/Caregiver Checklist used a combined sample of both typically developing and clinically diagnosed children from 0 to 12 years of age (n = 432). Subscale score coefficients for ages 0 and 1 year ranged from .83 to .91, with the GD score at .97. Coefficients for ages 2 and 3 years ranged from .79 to .96, with the GD score at .98 and .96, respectively. For ages 4 and 5 years, coefficients ranged from .80 to .96, with the GD score at .96 and .97, respectively.
2. Test-retest reliability: sixty-six parents from the standardization sample were interviewed a second time, with 13 to 18 days between administrations (average of two weeks). Correlation coefficients for subscale scores

ranged from .81 to .88, and the GD score was .92.

3. Inter-rater reliability: no information available.

Validity:

1. Content validity: Two exploratory common factor analyses—oblimin rotation and confirmatory factor analyses—indicated that items loaded primarily onto one main factor. Item response theory (Rasch model analyses) showed that the ranges of child ability and item difficulty for each scale were similar. For all scales, the range of person ability extends slightly below and slightly beyond the range of item difficulty, demonstrating that the items in all five scales dependably measure child development within the target skill range.
2. Concurrent validity: the Interview Form scales were compared to scales of similar constructs in the Vineland Adaptive Behavior Scales, Second Edition (Vineland II); Developmental Assessment of Young Children (DAYC); Peabody Developmental Motor scales, Second Edition (PDMS-2); and Preschool Language Scales, 4th Edition (PLS-4). All comparisons were conducted on subsets of the “clinical” sample ($n = 398$) of racially diverse children from 4 months to 12 years of age with pre-existing developmental, behavioral, or emotional disorders. The sample had twice as many boys (consistent with higher rates of developmental disabilities among boys in the sample). Correlations between scales of

similar constructs on the DP-3 subscales and Vineland II Adaptive Behavior Scales ranged from .68 (DP-3 Adaptive Behavior and Vineland II Daily Living Skills) to .85 (DP-3 Physical and Vineland II Motor Skills), and the Vineland II Adaptive Behavior Composite correlated with the DP-3 GD at .81.

Correlations between the same subscales on the DP-3 and DAYC ranged from .64 (Adaptive Behavior) to .71 (Communication), and the DAYC and DP-3 GD scores correlated at .72. The Communication scale on the DP-3 correlated with the PLS-4 Expressive Communication and Auditory Comprehension scales at .53 and .48, respectively. The DP-3 Physical scale correlated with PDMS-2, with coefficients of .56 for the PDMS-2 Grasping scale and .71 for the PDMS-2 Visual-Motor Integration scale.

The DP-3 Parent/Caregiver Checklist scores were compared to scores on the Vineland II and Adaptive Behavior Assessment System, Second Edition (ABAS-II) among parents of 99 typically developing and clinically diagnosed children. Correlations were similar to those of the Interview Form and the Vineland II, ranging from .61 (Vineland II Daily Living Skills and DP-3 Adaptive Behavior) to .78 (Vineland II Adaptive Behavior Composite and the DP-3 GD score). ABAS-II raw scores were available for 150 typically developing children, from 3 to 12 years of age, and correlations ranged from .45 (DP-3 Physical with ABAS-II Social) to .87

(DP-3 Cognitive with ABAS-II Functional Academics).

T-tests comparing the Interview Form and Parent/Caregiver Form standardization samples of children with samples of children with clinical problems (n = 398 and 56, respectively) showed significant differences for subscale and GD mean standard scores. Samples with clinical problems were further divided into groups, with one group of children with developmental delays, maternal drug use, mental retardation, and related disabilities and a second group with emotional, behavioral, and adjustment disorders. The second group also included children with speech and learning disabilities for Interview Form comparisons. As the author expected, the first group had lower scores across all scales and overall. Differences were significant among the Interview Form sample and non-significant among the Parent/Caregiver Form sample; the author attributed the non-significant differences to the small sample size.

3. Predictive validity: the original version of the Developmental Profile (1972) was used to predict 29 autistic children's functioning four to seven years later based on a questionnaire administered to mothers about their child's current level of functioning. Pearson product moment correlations between the DP and the parent's report were significant across all five scales, ranging from .43 to .61.

4. Construct validity: the author correlated scale scores with each other and with the composite GD score. Correlation coefficients ranged from .39 (Physical and Communication) to .79 (Cognitive and GD score). All scales had a higher correlation with the composite GD score than with any of the other scales.

Bias Analysis: No information available.

Method of Scoring: A clinician or a professional with training in child development, psychology, and/or education scores the results. The scorer records a "yes" or "no" for each question on each form. For hand scoring of the Interview Form, raw scores are calculated by counting one point for each "yes" and one point for each item below the basal, whereas the Parent/Caregiver raw score consists of "yes" responses. Points are subtotaled for each scale. The raw scale scores are converted into standard scores for each scale with the use of reference tables in the manual. To calculate the GD score, scale scores are first summed; the scorer then uses corresponding tables in the manual to identify a converted score. Additional tables in the manual indicate 95 percent confidence intervals for standard scores, descriptive categories (i.e., Delayed, Below Average, Average, Above Average, or Well Above Average), percentile ranks, and age equivalents. As an alternative to hand scoring, the scorer may manually enter responses into the computer for scoring with the WPS Test Report scoring software.

Interpretability: A Scoring/Profile form is used to interpret and convey test results to

parents. The scorer records raw scores (by scale only), standard scores, 95 percent confidence intervals, descriptive categories, percentile ranks, and age equivalents (by scale only) on the Scoring/Profile form. The scorer also plots on a graph standard scores for each scale and the GD score and records notes at the bottom of the page. The manual provides a detailed analysis of each scale if the clinician desires a more in-depth analysis of scale score patterns and/or an item analysis, instructions for integrating DP-3 results with information from other assessments and/or information obtained through remediation, and guidelines for interventions based on DP-3 results. The CD-ROM computer scoring program interprets scores with the following features: graphical representation, scale pattern analysis and scale-by-scale item analysis, and individualized intervention activities that address a child's weaknesses.

Training Support: No information available.

Adaptations/Special Instructions: No information available.

Report Preparation Support: The CD-ROM computer scoring program produces a clinician report and a parent report. The clinician report includes summaries of raw scores by scale,

standardized scores by scale, the confidence interval, the development range category, corresponding percentile, and age equivalent. It also provides a graphical depiction of the child's scores compared to standard score ranges as well as a scale interpretation and a list of suggested activities to develop each skill set. The parent report contains a description of the score range and descriptive category obtained for each scale and what it means to perform at each category level (Well Above Average, Above Average, Average, Below Average, or Delayed). The parent report also includes suggested home and family activities to develop skill sets.

References:

Alpern, Gerald D. *Developmental Profile 3 DP-3 Manual*. Los Angeles: Western Psychological Services, 2007.

Alpern, Gerald D. *Unlimited-use Scoring and Interpretation CD*. Los Angeles: Western Psychological Services, 2007.

Flanagan, Rosemary and Carlen Henington. "Review of Developmental Profile 3." In *The Eighteenth Mental Measurements Yearbook*, edited by Robert A. Spies, Janet F. Carlson, and Kurt F. Geisinger. Lincoln, NE: Buros Institute of Mental Measurements, 2010.

DEVEREUX EARLY CHILDHOOD ASSESSMENT FOR INFANTS AND TODDLERS (DECA-I/T), 2007 (2011 Update)

Authors:

Mary Mackrain, Paul LeBuffe, and Gregg Powell

Publisher:

Kaplan Early Learning Company
(800) 334-2014
<http://www.kaplanco.com>

Instrument:

<http://www.kaplanco.com/store/trans/productDetailForm.asp?CatID=5|LT1045|0&CollID=37615>

Initial Material Cost:

DECA-I/T Kit includes 50 record forms (20 Infant and 30 Toddler), Strategies Guide, User Guide, Set of 5 reproducible Parent Teacher Masters, 20 Parent Strategy Guides (For Now & Forever), 3 Adult Resilience Journals (Building Your Bounce), and Forms CD: \$199.95

Representativeness of Norming Sample:

The standardization sample consists of 2,183 infants and toddlers between 4 weeks and 3 years of age (45 percent infants and 55 percent toddlers).¹ The sample closely represents the 2006 *Statistical Abstract of the United States* in terms of characteristics such as gender, region, race/ethnicity, and socioeconomic status.

Languages:

English

¹ Infant was defined as age 4 weeks to 18 months and toddler as 18 months to 36 months.

Type of Assessment:

Observation and parent/caregiver report

Age Range and Administration Interval:

4 weeks to 36 months, administered three times per year, with at least a 4-week interval between assessments by the same adult.

Personnel, Training, Administration, and Scoring

Requirements: Administered by parents, adult family members, or teachers who are able to read at a grade 6 level and have observed the child's behavior for a minimum of 2 hours per day, 2 days each week, over a period of 4 weeks.

It takes 20 minutes to complete. Highly trained program members score, interpret, and translate standardized assessment instruments, standardized scores, and profiles. The DECA-I/T Scoring Program CD-ROM (\$99.95) may also be used to score the assessment. It is sold separately from the kit.

2-day DECA-I/T Implementation Training: \$270 (early bird rate) or \$300 (regular rate).

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (mostly .65 or higher).

Validity: 3 (all .5 or higher for concurrent)

Norming Sample Characteristics: 3 (normed within the past 15 years and nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Devereaux Early Childhood Assessment for Infants and Toddlers (DECA-I/T) is an observation used by parents or child care staff. It measures protective factors, or characteristics of the individual or environment thought to temper the negative effects of stress and lead to positive behavioral and psychological

outcomes in at-risk infants and toddlers age 1 year to 36 months. The DECA-I/T may also screen for risk in the social and emotional development of infants and toddlers. The instrument's two forms are designed for infants 1 to 18 months of age (DECA-I, 33 items) and toddlers 18 to 36 months of age (DECA-T, 36 items). The DECA-T is

modeled after the DECA (1999), which targets 2- to 5-year-olds.

The DECA-I has two protective factor scales (Initiative and Attachment/Relationships), and the DECA-T has three protective factor scales (Attachment/Relationships, Initiative, and Self-Regulation). The Initiative scale assesses the infant's or toddler's ability to use independent thought and actions (18 items on the DECA-I and 11 items on the DECA-T), and the Attachment/Relationship scale assesses the relationship between the infant or toddler and significant adults (15 items on the DECA-I and 18 items on the DECA-T). The 7-item Self-Regulation scale on the DECA-T assesses the toddler's ability to gain control of and manage emotions and sustain focus and attention. Both the DECA-I and DECA-T have a Total Protective Factors scale, which is a composite across the scales for each assessment; the scale provides an overall indication of the strength of the infant's or toddler's protective factors.

Other Languages: The authors are developing a Spanish version of the DECA-I/T.

Uses of Information: The DECA-I/T provides a profile of an infant's or toddler's social and emotional strengths and helps users identify children who may be experiencing social and emotional challenges. It is possible to compare scores between different raters for the same child in order to understand different child behavior across different environments. Programs may use the DECA-I/T to meet standards for social and emotional development of infants and toddlers,

compare children's social and emotional health over time, and develop strength-based programs to foster children's healthy social and emotional growth. The DECA-I/T also provides researchers with a measure of self-protective factors.

Reliability:

1. Internal consistency: Cronbach's alpha coefficients for Total Protective Factors on the DECA-I ranged from .90 to .94 for parent raters (.80 to .92 for the individual scales) and .93 to .94 for teacher raters (.87 to .93 for the individual scales) across the four infant age groups (1 to 3 months; 3 to 6 months; 6 to 9 months; and 9 to 18 months). For the DECA-T (18 to 36 months), coefficients for Total Protective Factors were .94 for parent raters (.79 to .92 for the individual scales) and .95 for teacher raters (.83 to .94 for the individual scales).
2. Test-retest reliability: the administration interval between parent and teacher rater administrations ranged from one to three days. Coefficients for Total Protective Factors on the DECA-I was .91 for parents (.86 to .94 for the individual scales), .84 for teachers (.83 to .84 for individual scales), and .85 for both parents and teachers. On the DECA-T, coefficients were .99 for parents (.92 to .99 for individual scales), .91 for teachers (.72 to .98 for individual scales), and .97 for both parents and teachers.
3. Inter-rater reliability: inter-rater reliability of one pair of teachers and parent raters for

Total Protective Factors on the DECA-I was .68 for parents (.59 to .76 for the individual scales) and .72 for teachers (.64 to .71 for the individual scales). On the DECA-T, the coefficient for total protective factors was .70 for parents (.62 to .72 for the individual scales) and .74 for teachers (.66 to .71 for the individual scales).

Validity:

1. Content validity: Items selected for the DECA-I/T were based on an extensive review of the literature on resilience; on focus groups with parents, teachers, and infant and early childhood mental health professionals; and on a review of infant and toddler social and emotional instruments. Reviewers noted that the literature review and assessment of other instruments need to be updated (Clark and McLellan 2010). Factor analysis procedures were used to select the items for each of the protective scales, resulting in a two-factor solution for infants and a three-factor solution for toddlers. The National Advisory Team and the Devereux Early Childhood Initiative (DECI) Research Advisory Board devised a naming convention for the factors that led to the current protective scales in each assessment.
2. Concurrent validity: thirty-five toddlers age 2 to 3 years were assessed with the original DECA and the DECA-T. Correlations between the DECA and DECA-T ranged from .83 (Initiative) to .91 (Total Protective Factors).

Fifteen infants and 69 toddlers with identified emotional and behavioral problems were found to have significantly lower protective factor scale scores and significantly higher behavioral concern than a matched sample of children with no identified emotional and behavioral problems. The authors performed sensitivity analysis to investigate the proportion of children in the sample with identified emotional and behavioral problems who scored as Area of Need (see Interpretability section); they also performed specificity analysis to investigate the proportion of children in the community sample who scored as Typical or Strength. Analyses among infants showed that sensitivity for the identified sample ranged from 27 percent (Initiative) to 47 percent (Attachment/Relationships). Specificity was 87 percent for each scale. Among toddlers, sensitivity for the identified sample ranged from 41 percent (Attachment/Relationships) to 57 percent (Total Protective Factors). Specificity ranged from 80 percent (Attachment/Relationships) to 87 percent (Initiative).

3. Construct validity: the authors reported findings consistent with DECA-I/T's theoretical construct. For both low- and high-risk children, higher protective factor scale scores were associated with better social and emotional health outcomes than were lower scores.

Bias Analysis: Authors compared the mean scores between black and white infants and toddlers and between Hispanic and white infants and toddlers and concluded most mean score differences were negligible or small except for the Attachment/Relationships scale for Hispanics and whites, which had a medium mean score difference.

Methods of Scoring: Raters complete the Record Form by indicating how often they observed the child performing particular behaviors in the past four weeks. For each item, they place a checkmark next to one of the following descriptors: never, rarely, occasionally, frequently, or very frequently. The DECA-I/T may be scored by using a Scoring Program CD-ROM, or a trained professional may score the assessment by hand. Hand scoring involves transferring the rater's checkmarks to a separate page of the Record Form, where corresponding boxes have raw score values that correspond to each rating: never = 0, rarely = 1, occasionally = 2, frequently = 3, and very frequently = 4. The scale raw scores are calculated by adding the raw scores for all items in each scale. The raw scores are converted into t-scores and percentiles with the use of the Individual Child Profile, which graphically displays the child's score across the scales. Individual Child Profile forms differ for infants by age group (1 to 3 months, 3 to 6 months, 6 to 9 months, and 9 to 18 months); therefore, it is important to select the appropriate profile based on the child's age. The Individual Child Profile provides separate sections for parent and teacher

ratings. The User's Guide includes tables to help interpret the differences between scores on the scales by different raters and over time.

Interpretability: Raw scores are converted into percentile scores, t-scores, or normal curve equivalents for interpretation and to compare ratings to typical scale scores. T-scores on the DECA-I/T range from 28 to 72 and are classified as Area of Need (40 or below), Typical (41 to 59), or Strength (60 and above). Scores should be interpreted in the context of other infant and toddler scores and cultural and family background.

Training Support: Training support is available through the Devereux Early Childhood Initiative and includes introduction and basic implementation sessions as well as a train-the-trainer session.

Adaptations/Special Instructions for Individuals with Disabilities: No information available.

Report Preparation Support: Individual Profiles graphically display results from the DECA-I/T scales. Several examples of DECA-I/T Individual Profiles are presented and interpreted, along with suggestions for an intervention plan.

References:

Clark, Jean N. "Review of the Devereux Early Childhood Assessment for Infants and Toddlers." In *The Eighteenth Mental Measurements Yearbook*, edited by Robert A. Spies, Janet F.

Carlson, and Kurt F. Geisinger. Lincoln, NE: The Buros Institute of Mental Measurements, 2010.

Mackrain, Mary, and Paul LeBuffe. *DECA-I/T CD Scoring Assistant*. Lewisville, NC: Kaplan Early Learning Company, 2007.

Mackrain, Mary, Paul LeBuffe, and Gregg Powell. *Devereux Early Childhood Assessment for Infants and Toddlers User's Guide*. Lewisville, NC: Kaplan Early Learning Company, 2007.

McLellan, Mary J. "Review of the Devereux Early Childhood Assessment for Infants and Toddlers." In *The Eighteenth Mental Measurements Yearbook*, edited by Robert A.

Spies, Janet F. Carlson, and Kurt F. Geisinger. Lincoln, NE: The Buros Institute of Mental Measurements, 2010.

Powell, Gregg, Mary Mackrain, and Paul LeBuffe. *Devereux Early Childhood Assessment for Infants and Toddlers: Technical Manual*. Lewisville, NC: Kaplan Early Learning Corporation, 2007.

The Devereux Early Childhood Initiative. "Building Skills for School Success." Available at [http://www.devereux.org/site/PageServer?pagename=deci_index].

EARLY COMMUNICATION INDICATOR (ECI), 2011 (2011 Update)

Authors:

Gayle J. Luze, Deborah L. Linebarger, Charles R. Greenwood, Judith J. Carta, Dale Walker, Carol Leitschuh, and Jane B. Atwater

Publisher:

Juniper Gardens Children's Project
(913) 321-3143
<http://www.igdi.ku.edu>

Instrument:

http://www.igdi.ku.edu/measures/ECI_Measures/

Initial Material Cost: Free; the publisher's web site includes administration and coding guidelines, practice videos, certification videos, and scoring forms.

Child Data System: \$1 per child annually for early childhood education programs.

Representativeness of Norming Sample: The norming study was conducted in 2010 with 5,883 children 6 through 36 months old from Early Head Start programs in two Midwestern states (Greenwood et al. 2010). The children lived in rural, urban, and suburban settings and were diverse with respect to race and ethnicity. Home languages were primarily English and Spanish. The sample also included children with special needs.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

Birth through 3 years old. Administered every 3 months.

Personnel, Training, Administration, and Scoring

Requirements: Administered by highly trained and certified program staff members or researchers.

Observers may be trained and certified in 4 to 6 hours to code the key skill communication elements, with each video taking about 30 minutes to code on average. Observers must attain at least 85 percent agreement (assessed annually) on an Observer Agreement Sheet that calculates inter-observer reliability. Within an organization, certified video coders may train others.

Assessors (adult play partners) become certified by scoring at least 81 percent on the Administration Checklist (an 18-item checklist covering set up, administration, and ending the session) based on the observer's rating.

Child observations last for 6 minutes.

Training materials for administering the ECI are available online for free and include training workshop PowerPoint slides, administration and scoring guidelines, and video demonstrations. Juniper Gardens Children's Project researchers are also available for in-person training (cost available upon request).

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (all .65 or higher)

Validity: 3 (all .5 or higher for concurrent)

Norming Sample Characteristics: 3 (normed within past 15 years and representative of children in Early Head Start programs)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Early Communication Indicator (ECI) is an observation of communicative interactions with a child and a

trained adult play partner. The ECI is part of a set of Individual Growth and Development Indicators (IGDI) for use by childcare practitioners to

monitor children's growth and development in key developmental areas. The other IGDIs measure social development, movement, problem solving, and parent-child interaction. The ECI is a play-based measure designed to elicit expressive communication among children from birth through 3 years old. A trained familiar adult (play partner) interacts with the child by using the standard ECI stimuli, including Fisher Price's *Little People Animal Farm* or *Little People Happy Home*. The farm includes a *Little People* farmer, cow, sheep, horse, pig, chicken, rooster, and food cart. The doll house includes three *Little People*, a bed, table, two chairs, stroller, and play center. For ECI assessments, the developers recommend using the toys without sound. The play partner prepares for the assessment by following the administration instructions that detail the set up, selection of toys, and end of the session. The play partner encourages and supports the child's communication without being directive. Observations of the play session may be live or videotaped. Using the ECI Recording Sheet, observers document how often four skill elements occur in one-minute intervals across the assessment's six minutes. Skill elements include Gestures, Vocalizations, Single-Word Utterances, and Multiple-Word Utterances; together, these elements make up a Total Communication Score. The observer records the frequency of occurrence of each skill type in one-minute intervals.

Other Languages: None.

Uses of Information: Data from the ECI may be used to screen children's expressive

communication abilities, monitor children's expressive communication progress, inform interventions, and educate parents. The data may also be used for program improvement and research purposes.

Reliability:

1. Internal consistency reliability: split-half reliability coefficients for the mean of all the child's total communication scores was .89 (Luze et al. 2001). The convenience sample included 50 infants and toddlers selected from five childcare centers in two cities; the infants and toddlers varied in race, gender, socioeconomic status, disability status, and mother's education level.
2. Test-retest reliability: analysis was conducted on 25 children from the pilot study who completed nine monthly ECI assessments. Test-retest coefficients (type not specified) for total communication was .89 (Greenwood et al. 2006).
3. Inter-rater reliability: data are based on a composite sample from the study conducted by Luze et al (2001) and on pooled samples from three additional studies (n = 1,653) that included toddlers from 36 center- and home-based programs in Kansas; the toddlers were racially diverse and bilingual and came from predominantly low-income families (http://www.igdi.ku.edu/measures/ECI_Measures/DetailsofECInormativestudysamples.htm). Data from 10 percent of the completed assessments were randomly selected to test inter-rater reliability. Two trained assessors

coded each video. Average agreement was 90 percent for overall communication and 81 percent for gestures, 80 percent for vocalizations, 70 percent for single words, and 72 percent for multiple words (Greenwood et al. 2010).

Validity:

1. Content validity: Luze et al. (2001) conducted a review of the communication and language development literature and of existing instruments to identify the four ECI skill elements as well as Social Attention (i.e., intentional and sustained gaze toward the play partner) and toys used in the observation (Fisher Price's *Little People Animal Farm* or *Little People Happy Home*). Authors pilot tested 25 children from 5 through 36 months old from three child care programs. Social Attention was dropped from the ECI because it did not increase with age as expected.
2. Concurrent validity: the ECI correlated with the Expressive Language subscale of the Preschool Language Scale at .62 and with the Caregiver Communication Measure at .51 (Luze et al. 2001).
3. Construct validity: results with the norming sample suggest that the ECI is sensitive to the change in onset of the key skill elements, pattern of growth, and outcomes for key skill elements at 36 months of age.

Bias Analysis: Growth curve analyses indicated that children's communication growth is not influenced by gender or the language spoken at home (Greenwood et al. 2010).

Methods of Scoring: The observer adds the number of times that each skill element-Gestures, Vocalizations, and Single- and Multiple-Word Utterances-is observed across the six minutes and records the information on an ECI Recording Sheet. Total communication, the sum of the individual frequencies, is scored within and across skill elements. Each skill element is weighted differently when determining the Total Communication Score; that is, Single- and Multiple- Word Utterances receive a weight of two and three, respectively. Weights approximate an absolute estimate of the child's total words during the assessment and are automatically applied when raw scores are entered into the Child Data System (the developer's online data entry system). Alternatively, raw scores may be data entered, with the weights applied during analysis.

Interpretability: No information available.

Training Support: Training materials are available online and include administration guidelines; coding definitions; PowerPoint presentations from training workshops covering background, administration, scoring, and training of others; and videos to practice and code observations. The Juniper Gardens Children's Project ECI researchers are also available for in-person training and provide ongoing training support as necessary.

Adaptations/Special Instructions for Individuals with Disabilities: Gestures and Vocalizations may be used with children whose speech is limited by developmental delays or

disabilities. For children with physical disabilities, the authors recommend modifying the toys' placement (e.g., move the toys closer, prop the child up to facilitate access to the toys). Adults may also move toys upon the request of the child (e.g., the child wants the horse to jump over the fence but cannot pick up the horse). For children with visual impairments, the play partner may identify the toy, orient the child to the toy, and allow the child to touch each toy to become familiar with it. The play partner may position him-/herself in front of children with hearing impairments-instead of sitting next to the child-with the toys placed between them in order to facilitate sign language and lip reading. For children with hearing aids or audio tuner, the play partner should choose the optimal position that enables the child to see and hear him/her.

Report Preparation Support: ECI data entered into the Child Data System may be used to generate several types of progress reports, including reports for individual children, programs, or several programs. Office of Special Education Programs (OSEP) reports may also be generated to share information with parents, to monitor progress or program improvement, or to meet reporting requirements.

References:

Carta, Judith, Charles Greenwood, Dale Walker, and Jay Buzhardt. *Using IGDI: Monitoring Progress and Improving Intervention for Infants and Young Children*. Baltimore: Paul H. Brookes Publishing Co., Inc., 2010.

Greenwood, Charles R., Judith Carta, Dale Walker, Kere Hughes, and Mary Weathers. "Preliminary Investigations of the Application of the Early Communication Indicator (ECI) for Infants and Toddlers." *Journal of Early Intervention*, vol. 28, no. 3, 2006, pp. 178-196.

Greenwood, Charles R., Dale Walker, and Jay Buzhardt. "The Early Communication Indicator for Infants and Toddlers: Early Head Start Growth Norms from Two States." *Journal of Early Intervention*, vol. 32, no. 5, 2010, pp. 310-334.

Juniper Gardens Children's Project. "Early Communication Indicator (ECI): Administration Guidelines." Available at [http://www.igdi.ku.edu/training/ECI_training/ECI_Administration_Instructions-2005.pdf]. 2005.

Juniper Gardens Children's Project. "ECI Certification Videos." Available at [http://www.igdi.ku.edu/training/ECI_training/ECI_videos/index.htm]. 2009.

Juniper Gardens Children's Project. "IGDI Child Data System." Available at [http://www.igdi.ku.edu/data_system/index.htm]. 2009.

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

EARLY COPING INVENTORY (ECI), 1988

Authors:

Shirley Zeitlin, G. Gordon Williamson, and Margery Szczepanski

Publisher:

Scholastic Testing Service, Inc.
www.ststesting.com
1-800-642-6787

Initial Material Cost:

ECI Manual and 20 forms: \$51

Representativeness of Norming Sample:

None described.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

Children whose developmental age is between 4 and 36 months

Personnel, Training, Administration, and Scoring Requirements:

Nonprofessionals who are knowledgeable in infant development may administer and score the inventory. If observers are not familiar with the child, they should observe the child at least 3 times in different situations. Interpretation should be done by a professional with a background in early development and behavioral sciences. Administration time varies depending on the observer's familiarity with the child.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: Interrater: 3 (.65 or higher)

Validity: 1 (content validity only reported)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (may be administered by a nonprofessional but scored by a highly trained individual).

Description: The Early Coping Inventory (ECI) is an observation instrument used for assessing the coping-related behavior of children whose chronological or developmental age is between 4 and 36 months. The ECI's 48 items are divided into 3 coping clusters: Sensorimotor Organization, Reactive Behavior, and Self-Initiated Behavior. Each item is rated on a five-point scale ranging from ineffective coping (1) to consistently effective coping across situations (5).

Uses of Information: Analysis of a child's scores on the instrument provides information about level of coping, style, and specific strengths and weaknesses. The findings can then be used to create educational and therapeutic interventions.

In addition, the ECI can be used to involve parents in its use as a means of increasing knowledge of the child and communication with staff. The ECI can also be used to support staff development and training to increase observation skills, expand their domain of concern, facilitate teamwork, and measure child progress. The manual provides a chapter on how the ECI can be used for these purposes.

Reliability:¹ Reliability was established with a group of observers who completed the ECI after viewing videotapes of four young children twice,

¹ The reliability tests were conducted using the research version of the ECI, which had 54 items instead of 48 items in the current version.

with a 6-week interval between viewing.

(1) Interrater reliability (using Guildford's formula): At the first viewing, the reliability coefficients for the three coping clusters and the adaptive behavioral index ranged from .80 (sensorimotor organization) to .94 (self-initiative behavior) and at the second viewing from .87 (sensorimotor organization) to .93 (self-initiative behavior). The authors also tested for the level of agreement between the observers' scores with the scores of an expert panel for each of the ECI items. The "concordance index" showed that the agreements in the item scores within each coping cluster ranged from a mean of 41 percent (reactive behavior) to a mean of 52 percent (sensorimotor organization). (2) Test-retest reliability (six-week interval): Friedman's analysis of variance test was used to test for significant differences between ECI test-retest scores for each child on the coping clusters and the adaptive behavioral index. The authors reported no statistical significant shift in scoring on 11 of the 16 tests.

Validity: (1) Content validity: Item content and definitions of coping constructs were primarily derived from a review of the early childhood, coping-related literature. A panel of six judges then reviewed the selected items for fidelity to the coping constructs. The EIC was then administered to three different samples and the responses were factor analyzed. The results of the factor analysis provided the basis for creating the three coping clusters.

Method of Scoring: Raw score totals are calculated for sensorimotor organization, reactive

behavior, and self-initiated behavior by summing the items ratings scale numeric values. A table is used to convert the raw scores into Effectiveness scores, which can be plotted on the Coping Profile and used to compare the child's level of effectiveness in the three categories. Another table converts the sum of the effectiveness scores into an Adaptive Behavior Index score. A list of six to eight of the Most and Least Adaptive Coping Behaviors is also compiled to aid intervention planning.

Interpretability: Higher scores indicate the use of more effective coping behaviors in adapting to stresses in everyday living. The Adaptive Behavior Index indicates the child's general level of effectiveness in using adaptive behaviors to cope and whether or not intervention is needed. The Coping Profile, which graphically displays the effectiveness scores, shows strengths and vulnerabilities in coping behavior. The authors recommend taking advantage of areas of strengths when planning intervention activities to increase competence in the weaker areas. In the same way, the list of the Most and Least Adaptive Coping Behaviors can also be used to develop the intervention plan. The manual provides a table that contains a descriptive interpretation for the Effectiveness and Adaptive Behavior Index numeric scores and several case studies to illustrate how to interpret the results.

Training Support: Two examples are provided to illustrate use of the rating scale and three case studies are given on how to interpret the results. Appendix D of the manual also

contains case studies on how to develop intervention plans.

Adaptations/Special Instructions for Individuals with Disabilities: Instructions are given in the manual on how to rate a child with a disability.

Report Preparation Support: None.

References:

Zeitlin, Shirley, G. Gordon Williamson, and Margery Szczepanski. Early Coping Inventory: A Measure of Adaptive Behavior. Bensenville, Illinois: Scholastic Testing Service, 1988.

EARLY HEAD START RESEARCH AND EVALUATION PROJECT (EHSREP) PARENT INTERVIEWS AND CHILD ASSESSMENTS, 1996 - 2001

Authors:

John Love and other project staff, in collaboration with the Early Head Start Research Consortium

Publisher:

Mathematica Policy Research, Inc.
Contact Publications, 609-275-2350,
jallen@mathematica-mpr.com.

Lists of measures and copies of the interviews can be found at www.mathematica-mpr.com or http://www.acf.hhs.gov/programs/opre/ehs/ehs_research/index.html.

Initial Material Cost:

None

Representativeness of Norming Sample:

None described

Languages:

English, Spanish

Type of Assessment:

Parent report, observation, and direct child assessments

Age Range and Administration Interval:

For child-related questions and assessments, prenatal through 36 months.

For parent-related questions, all adults.

Personnel, Training, Administration, and Scoring Requirements:

The requirements vary, depending on the particular measure. Some may be completed by parents and scored by a person with minimal training. Others must be administered and scored by a highly trained individual.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (.65 and higher)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (some measures require administration and scoring by a highly trained individual)

Description: The Parent Interviews (PIs), Father Interviews (FI), and Child Assessments (CA) developed for the national Early Head Start Research and Evaluation Project (EHSRE) were designed to include instruments that assess potential program effects on a variety of domains. The instruments included in each data collection interview/assessment were drawn from a variety of sources and include published instruments (many of them are described in individual entries in this resource guide), questions drawn from other large national surveys, and questions that were developed specifically for this study.

The assessments used in the evaluation are summarized in table format at <http://www.acf.hhs.gov/programs/opre/research/project/early-head-start-research-and-evaluation-project-ehsre-1996-2010>. The table in Appendix C lists key child and family measures selected for the evaluation. In addition to scales and standardized tests, the interviews and assessments included a number of single items that are simple to administer and use for comparison with the national evaluation results. These include questions about bedtime routines, reading to children at bedtime, frequency of reading to

children, and spanking. The interviews can be found at the web addresses listed above.

The results of the evaluation (through age 3) are included in two reports and their appendices and are available at www.mathematica-mpr.com and http://www.acf.hhs.gov/programs/opre/ehs/ehs_research/index.html.

Uses of Information: The Early Head Start interviews and assessments can be used by programs to obtain a wide range of parent, child, and service use information useful for performance measurement that can be compared to the national study findings. For copyrighted instruments, programs must obtain permission to use the assessments and must pay for their use.

Reliability: The technical appendices of the two reports include internal consistency reliability for all of the summary scores. As a general rule, summary scores were not included in the report if their reliability was not above .65. The single item questions do not require computing summary scores.

Validity: The assessments were included in the evaluation because they had been used before in large studies and had demonstrated construct validity. Validity work based on the data collected was not reported in the two reports.

Method of Scoring: Each assessment is scored according to the rules and advice from the assessment developers or publishers. Some required complex computer scoring programs and others could be scored by hand and were a simple

percentage. Scoring procedures for each measure are summarized in the reports. The single item questions do not require computing summary scores.

Interpretability: Some of the assessments were easily interpretable, while others required a well-trained individual.

Training Support: As part of the evaluation project, in-depth training manuals were developed; these can be obtained by requesting them from Jackie Allen at jallen@mathematica-mpr.com. In addition to the manuals, interviewers and assessors attended a central training session and had to meet rigorous standards before administering the study instruments. Mathematica is not providing any training support for the measures.

Adaptations/Special Instructions for Individuals with Disabilities: Contact Mathematica for more information about how the protocols were adapted for use with individuals with disabilities.

Report Preparation Support: None described.

References:

ACYF. "Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families. Volume I: Technical Report." Washington, DC: Administration on Children, Youth and Families, 2001. www.mathematica-mpr.com or http://www.acf.hhs.gov/programs/opre/ehs/ehs_research/index.html.

ACF. "Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start." Washington, DC: Administration for Children and Families, June 2002. www.mathematica-mpr.com or http://www.acf.hhs.gov/programs/opre/ehs/ehs_research/index.html.

Sprachman, S., C. DeSaw, L. Mendenko, M. Salem, K. Boller, and B. Kolln. "Early Head Start National Evaluation Data Collection Training

Manual for 6-Month Parents Services Interview, 14-Month Parent-Child Interview and Assessment, and Child Care Observations." Princeton, NJ: Mathematica Policy Research, Inc., December 1996.

<http://www.mathematica-mpr.com/>

EARLY LEARNING ACCOMPLISHMENT PROFILE REVISED EDITION (E-LAP), 1995

Authors:

M. Elayne Glover, Jodi L. Preminger, Anne R. Sanford

Publisher:

Kaplan Press
(800) 334-2014

Initial Material Cost:

Early Learning Accomplishment Profile (E-LAP) Kit: \$335: Includes E-LAP Scoring Booklets, E-LAP Manual, and E-LAP Activity Cards, and Content and Overview Video.

Demonstration Video (training) costs \$66.

Representativeness of Norming Sample:

Research sample was representative of Year 2000 based on the 1995 US Census Bureau population projection. Sample included 285 children ages 2 to 44 months old, including children with typical and atypical development. A stratified sampling procedure was used based on geographic region, age, race, gender, and type of setting. Children were selected from child care centers, Early Head Start programs, and individual homes. The sample included a representative percentage of children with disabilities (reflecting the U.S. rates for children with disabilities – U.S. Census Bureau, 1995). The assessments were conducted between November 1999 and July 2000.

Languages:

English and Spanish

Type of Assessment:

Observation

Age Range and Administration Interval:

0 to 36 months

Personnel, Training, Administration, and Scoring Requirements:

In addition to having knowledge about the target population, a 2-day training workshop is suggested prior to using the E-LAP. It takes an experienced examiner at least an hour to administer the E-LAP and approximately 10 minutes per domain to score

Summary:

Initial Material Cost: 3 (> \$200)

Reliability: 3 (.65 or higher for internal consistency, test-retest, and interrater reliability)

Validity: 3 (.5 or higher for concurrent validity)

Norming Sample Characteristics: 3 (nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The E-LAP is designed to help assess overall development of children with special needs functioning, based on developmental milestones focusing on the birth through 3-year age range. The E-LAP is a criterion-referenced tool that focuses on the following domains: Gross Motor, Fine Motor, Cognitive, Language, Self-Help, and Social/Emotional skills. Items in the E-LAP were drawn from various early childhood assessment instruments.

Uses of Information: The E-LAP provides guidance to early childhood programs in assessment and programming for infants, young children, and children with special needs.

Reliability: (1) Internal consistency reliability (Cronbach's alpha) ranges from .84 to .98 for Gross Motor, with a total of .99; .90 to .96 for Fine Motor, with a total of .98; .96 to .97 for Cognitive, with a total of .99; .91 to .96 for Language, with a total of .98; .93 to .97 for Self-Help, with a total of

.98; and .87 to .91 for Social Emotional, with a total of .96. (2) Test-retest reliability was measured one to three weeks apart only for a small subsample of children (92). Correlations for the domains ranged from .96 to .99. (3) Inter-rater reliability correlation coefficients for the domains ranged from .96 to .99.

Validity: (1) Concurrent validity was tested using the Mental and Motor Scales of the BSID-II (Bayley Scales of Infant Development). Results indicate a strong correlation (.90 to .97) between the E-LAP and BSID-II scored in each domain for the overall sample. In the 2- to 12-month sample, correlations ranged from .83 to .95, and in the 13- to 24-month sample, correlations ranged from .72 to .88. Correlations in the 25- to 36-month sample were lower, ranging from .47 to .83. The manual mentions that this lower correlation, in addition to other analytical information, may indicate that the ELAP is a less effective assessment tool for older children.

Method of Scoring: The examiner must first calculate the child's chronological age (by months) to determine the appropriate starting point in each domain. Each item is marked with a plus (+) if the child exhibits the criterion-referenced behavior or a minus (-) if the skill is not demonstrated by the child. Examiners must establish a basal (8 consecutive items successfully completed) and a ceiling (3 errors out of 5 consecutive items). All items prior to the basal are counted as correct. The raw score represents these items plus the number of items successfully completed in the domain up to the ceiling. The

manual provides further guidelines for computing the raw score for the domain, and calculating the corresponding developmental age-range that the child falls within. It also suggests that any modifications of the procedures or use of adaptive equipment be included in the comments section of the scoring booklet in order to better understand a child's skills. The Content and Overview video also includes information on scoring the E-LAP.

Interpretability: E-LAP is a non-standardized test, and should be used in conjunction with norm-referenced assessments to determine whether or not a child has a disability. In computing scores, the manual also states that it is important for the examiner to be aware that the normative developmental age assigned to a specific item varies among research sources, and that developmental ages need to be viewed as approximate. Although the manual does not provide normed scores, it does provide means and standard deviations for the scores obtained in the various domains for the project sample (including both typical and atypical children) as well as the core sample (including only children with presumed typical development).

Training Support: There is a demonstration video available to assist transdisciplinary teams with using the E-LAP. There are also individuals from the Chapel Hill Training-Outreach Project available for training on the E-LAP. Information on training can be obtained by calling 800-334-2014, ext.5100.

Adaptations/Special Instructions for Individuals with Disabilities: The manual suggests that in the case of children with disabilities, reports of diagnostic results should be used to provide information regarding the child's developmental level of functioning (to help determine the appropriate point for beginning the assessment process). If that information is not available, the manual suggests that the assessor begin administering the E-LAP at half of the child's chronological age, which would allow for the establishment of the basal. The manual assumes that the teacher will create appropriate developmental milestones for children with more involved disabilities, and if necessary, should appropriately modify these milestones into sub-objectives for the child.

Report Preparation Support: The information in the E-LAP recording procedures suggest that the absence of certain skills (as demonstrated through the assessment) should be incorporated into the child's Individualized Family Service Plan (IFSP).

References:

Glover, Elayne M., Jodi L. Preminger, and Anne R. Sanford. *Early Learning Accomplishment Profile Revised Edition (E-LAP)*. Lewisville, NC: Kaplan Press, 1995.

Hardin, Belinda J., and Ellen S. Peisner-Feinberg. *The Early Learning Accomplishment Profile (Early LAP) Examiner's Manual and Reliability and Validity Technical Report*. Lewisville, NC: Kaplan Press, 2001.

EARLY LITERACY SKILLS ASSESSMENT (ELSA), 2005 (2011 Update)

Authors:

Andrea DeBruin-Parecki, Ph.D.

Publisher:

High/Scope Press

(734) 485-2000

<http://www.highscope.org>

Instrument:

<http://www.highscope.org/Content.asp?ContentId=114>

Initial Material Cost:

Complete kit: 2 copies of the book (*Violet's Adventure* or *Dante Grows Up*), 1 User Guide, 2 scoring pads (60 each, Score Sheet and Child Summary forms), 12 Class Summary forms, and 60 Family Report forms: \$149.95.

Representativeness of Norming Sample:

No norming sample

Languages:

English, Spanish

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

Ages 3 to 5, including kindergarteners. The developer recommends two administrations: once at the beginning of the school year and once at the end. Additional administrations should use the other of the two books (*Violet's Adventure* or *Dante Grows Up*).

Personnel, Training, Administration, and Scoring

Requirements: Administered by a teacher or other trained adult. The assessor should practice on a child ahead of time to become familiar with the story and questions. The publisher offers optional one- or two-day workshops on how to administer the assessment. Two-day workshop fees-Option 1: High/Scope contracts with an agency to train up to 25 teachers for a flat fee of \$4,300

Option 2: Tuition of \$225 per participant

Option 3: For up to 15 teachers, daily fee of \$1,200 plus the High/Scope consultant's travel expenses

One-day workshops-Hosted by High/Scope at variable cost (schedule at

<http://www.highscope.org/Content.asp?ContentId=406>)

Scoring tutorial DVD available for purchase for \$49.95.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (mostly .65 or higher).

Validity: 2 (mostly .5 or below for concurrent)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 2 (administered and scored by someone with basic familiarity with the test, usually a preschool teacher)

Description: The Early Literacy Skills Assessment (ELSA) is an individually administered assessment for 3- to 5-year-olds that is designed to measure pre-literacy skills. The assessment is administered through a shared reading exercise, with questions asked about the story book used (*Violet's Adventure* or *Dante Grows Up*). The ELSA consists of 23 items, with the following four subtests: Comprehension,

Phonological Awareness, Alphabetic Principle, and Concepts About Print. The Comprehension skills subtest asks the child to guess what will happen next in the story, retell in order what happened at the end of the story, and make a personal connection to the story by relating characters or plot to real people or experiences. The Phonological Awareness subtest measures rhyming, segmentation, and phonemic awareness.

It asks the child to match words with similar ending sounds, orally divide words into syllable chunks, and identify and match initial sounds. The Alphabetic Principle subtest assesses "sense of word" by asking the child to identify a given word and indicate the first and last letters in the word. Concepts About Print are assessed through questions on how to hold the book, where the reader begins, and what direction to read the text.

Other Languages: Spanish storybook versions of the ELSA, *La Aventura de Violeta* and *El Cambio En Dante*, perform similarly to the English storybooks based on factor analysis, suggesting that it is feasible to use the ELSA to study the emergent literacy skills of Spanish-speaking children and compare skills across distinct populations (Cheadle 2007). The factor analysis was conducted on a sample of 307 Spanish-speaking preschool children whose average age was 4.5 years. The sample was half female, 10 percent had special needs, and 80 percent had limited English proficiency. Results of the factor analysis supported the constructs for each subtest based on a variety of confirmatory factor analysis statistical tests, including the Chi Square Difference test, Comparative Fit Index, Tucker Lewis Index, and Root Mean Square Error of Approximation. The study reported high factor loadings and noted that items had adequate discrimination.

Uses of Information: The ELSA is used to measure preschoolers' early literacy skills and measure changes in these skills from the beginning to the end of the school year. Teachers

may use the results of the assessment to determine children's literacy levels and to inform instruction.

Reliability: The developer pilot-tested the ELSA on a convenience sample of 630 preschool children (average age of 4 years) from three preschool sites in Florida, Maine, and Michigan. Children were primarily African American and white, just over half of the sample was male, and one-fourth had special needs. Scores were analyzed for children in fall 2003 (n = 565) and in spring 2004 (n = 513). Cheadle (2007) also tested the ELSA on samples of English-speaking children. The *Dante Grows Up* sample consisted of 535 preschool children (average age of about 4.5 years), approximately half female, 6.2 percent special needs, and racially diverse (one-fourth Hispanic and black, respectively, and one-third white). The *Violet's Adventure* sample consisted of 505 preschoolers (average age approximately 4.5 years), half female, 3.7 percent special needs, and predominately white. For both groups, about one-third of the preschoolers were enrolled in Head Start programs.

1. Internal consistency reliability: the developer reported Cronbach's alphas by subtest and composite scores for fall and spring administrations. The composite score coefficients were .82 and .88 for the pre-test and post-test samples, respectively. Across subtests, coefficients ranged from .57 (Phonological Awareness) to .76 (Alphabetic Principle) for the pre-test sample and from .67 (Phonological Awareness) to .83 (Comprehension) for the post-test sample.

Cheadle (2007) analyzed scale scores with Cronbach's alpha and item response theory (IRT) reliability coefficients for both the Dante Grows Up and Violet's Adventure ELSA versions across the subtests and overall by pre-test and post-test periods. IRT reliability coefficients are reported here because scores were not normally distributed among the samples. For composite scores, pre-test and post-test IRT reliability estimates were .89 and .91, respectively. Across subtests, coefficients ranged from .81 to .96 for the pre-test sample and from .80 to .93 for the post-test sample.

2. Test-retest reliability: no information available.
3. Inter-rater reliability: no information available.

Validity:

1. Content validity: the developer asserts that content validity has been established by linking all items in the four subtests to key principles acknowledged in scientifically based literature. The developer conducted factor analysis based on data collected concurrently in the fall (for a subsample of 213 children) that supported the development of the four subtests in the ELSA. The Concepts About Print factor analysis initially showed two distinct factors, although changes were made to correct for the factors. Cheadle (2007) conducted extensive factor analysis and found high factor loadings that

demonstrated adequate item discrimination and consistent factor structures between pre- and post-testing.

2. Concurrent validity: the developer correlated subtests of the ELSA Violet's Adventure with the Get Ready to Read! (GRR) total score. The ELSA Phonological Awareness correlated with GRR at .53, the ELSA Alphabetic Principle and the GRR correlated at .64, and the ELSA Concepts About Print correlated with the GRR at .43. The total of the three subtests from ELSA correlated with the GRR total at .67. Cheadle (2007) correlated ELSA subtests with the Woodcock Johnson-III (WJ-III) (test not specified) and the Preschool Comprehensive Test of Phonological Processing (Pre-CTOPP). Coefficients were .74 for the WJ-III and the ELSA Alphabetic Principal and 1.0, .64, and .65 for the Pre-CTOPP and the ELSA Alphabetic Principal, Phonological Awareness, and Concepts About Print subtests, respectively.

With respect to subgroup differences, the developer presents differences in mean ELSA scores by age and by disability status for each subtest. Chronological age was positively related to ELSA performance on the four subtests such that raw score means increased with age. For scores by disability status, typical children outperformed their peers with disabilities on all ELSA subtests except post-test comprehension. Cheadle (2007) indicates the English and Spanish versions of the assessment have floor effects such that

versions inadequately distinguish younger children and children with weaker skills.

Cheadle recommends assessing and scoring children age 45 months or older to reduce floor effects, especially among lower-skilled, disadvantaged populations.

Methods of Scoring: The assessor scores the ELSA while reading the story and conducting the assessment. Raw scores are calculated by counting the number of correct responses to each item and then summing the scores by subtest and in total.

Interpretability: Score summary sheets are available at the child and class levels. The Child Summary sheet indicates three levels that correspond to different raw score ranges for each subtest. The manual notes that the levels best correspond to activities in High/Scope's Growing Readers Early Literacy Curriculum, which provides differentiated early literacy lessons and activities for each content area and level. On the Class Summary sheet, the assessor enters each child's levels by subtest to plan leveled literacy activities and lessons.

Training Support: The ELSA may be administered by a preschool teacher or other trained adult. The developer suggests one practice administration. The publisher offers optional one- or two-day workshops on how to administer the assessment. The one-day workshop is recommended in conjunction with the publisher's Growing Readers Early Literacy Curriculum two-day workshop. A scoring tutorial DVD is available for purchase online, and viewing is recommended

to establish reliability and consistency across those scoring the assessment.

Adaptations/Special Instructions for Individuals with Disabilities: The ELSA may be administered in two sittings for children with special needs and/or English language learners. The assessor marks where in the story the child stopped and then uses the pictures to summarize the story up to that point before commencing with the second half of the story.

Report Preparation Support: Information from the Child Summary sheet may be used to fill in the About Your Child as a Reader sheet, a tool to help interpret children's scores for parents. The manual provides scripted comments describing children's skills in English and Spanish for each of the three levels and by subtest; however, sharing actual scores or levels with parents is discouraged.

References:

Cheadle, Jacob E. "The Early Literacy Skills Assessment (ELSA) Psychometric Report for Both English and Spanish Versions." Ypsilanti, MI: High/Scope Early Childhood Reading Institute, 2007.

DeBruin-Parecki, Andrea. *Early Literacy Skills Assessment User Guide*. Ypsilanti, MI: High/Scope Press, 2005.

DeBruin-Parecki, Andrea. *ELSA (Spanish) Featuring El Cambio En Dante*. Ypsilanti, MI: High/Scope Press, 2005.

DeBruin-Parecki, Andrea. *ELSA (Spanish) Featuring La Aventura De Violeta*. Ypsilanti, MI: High/Scope Press, 2005.

Geisinger, K.F., and Linda L. Murphy. "Early Literacy Skills Assessment." In *The Seventeenth Mental Measurements Yearbook*, edited by Kurt F.

Geisinger, Robert A. Spies, Janet F. Carlson, and Barbara S. Plake. Lincoln, NE: Buros Institute of Mental Measurements, 2007.

Scoring the ELSA: Establishing Reliability DVD. Ypsilanti, MI: High/Scope, 2005.

EXPRESSIVE ONE-WORD PICTURE VOCABULARY TEST-4 (EOWPVT-4), 2010 (2011 Update)

Authors:

Nancy Martin and Rick Brownell

Publisher:

Academic Therapy Publications
(800) 422-7249
<http://www.academictherapy.com>

Instrument:

<http://www.academictherapy.com/detailATP.tpl?action=search&cart=%5bcart%5d&eqskudatarq=8543-0&eqTitledatarq=Expressive%20One-Word%20Picture%20Vocabulary%20Test%20-%204th%20Edition%20%28EOWPVT-4%29&eqvendordatarq=ATP&bobby=%5Bbobby%5D&bob=%5Bbob%5D&TBL=%5bt>

Initial Material Cost:

The EOWPVT-4 English Edition Test Kit with a manual, test plates, and 25 record forms, in a vinyl folder: \$175.

Representativeness of Norming Sample:

The norms were based on a nationally representative sample of 2,394 individuals age 2 through 80+ years (highest age was 103), selected to match the population distribution of the U.S. Census. Testing was conducted at 84 sites in 26 states. The sample was stratified by region, metropolitan area, ethnicity, gender, and education level. The Spanish speaking sample is described below (see Other Languages).

Languages:

English, Spanish

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

2 years to 80+ years

Personnel, Training, Administration, and Scoring

Requirements: Administered by a highly trained program staff member. Assessors need to be professionally trained in assessing cognitive functions and planning remediation activities; they are typically school psychologists, speech pathologists, counselors, or rehabilitation specialists. The assessor should conduct several practice trials before administering the assessment. Individuals interpreting the results must be formally trained in psychometrics and the use of derived scores.

The test is untimed and may be administered in 20 minutes. Scoring takes about 5 minutes.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (all .65 or higher)

Validity: 1 (none described)

Norming Sample Characteristics: 3 (normed nationally within past 15 years)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Expressive One-Word Picture Vocabulary Test-4 (EOWPVT-4) is an individually administered adaptive test that measures English expressive vocabulary normed for age 2 through 80+ years. Assessors present color pictures one by one to a child; the child identifies the pictures orally. The measure includes 190 items ordered by increasing

difficulty, with a basal of eight consecutive correct responses and a ceiling of six consecutive incorrect responses. The EOWPVT-4 was updated from its previous version to include national norms for adults age 80 years and older (the earlier version was normed for adults up to 18 years, 11 months) and contains additional items to reflect the expanded age range. The first half of the

items are sequenced with difficulty indices for children age 2 through 12 years while the second half of the items are sequenced with difficulty indices for children older than 12 years.

Other Languages: The EOWPVT-Spanish Bilingual Edition (SBE) was published in 2001 and normed on a national sample of Spanish-bilingual individuals 4 years old through 12 years, 11 months; 50 percent of the sample had mothers with less than a high school diploma (Brownell 2001). Record forms for the SBE include acceptable responses in both English and Spanish. The SBE basal and ceiling rules differ from the rules for the English Expressive One-Word Picture Vocabulary Test, 2000 Edition (EOWPVT).

Uses of Information: The EOWPVT-4 measures an individual's vocabulary, screens for early language delay in young children, identifies difficulties in reading or expressing words, assesses the English vocabulary of an English language learner, and, when administered alongside the Receptive One-Word Vocabulary Test (ROWPVT), measures expressive aphasia. The measure may also evaluate cognitive ability and intervention programs designed to increase vocabulary or improve reading comprehension.

Reliability:

1. Internal consistency reliability: Cronbach's alpha coefficients were calculated on norming sample participants and ranged from .94 to .95 for 2- to 5-year-olds.
2. Test-retest reliability: the sample consisted of 78 individuals evenly divided between genders, predominantly Caucasian and Hispanic, and varied in terms of educational achievement and geographic location (ages not reported). The correlations between the scores of two administrations (conducted with an average of 19 days between tests) were .98 for the raw score and .97 for the standard score.
3. Inter-rater reliability: not available for the EOWPVT-4. When two assessors administered the EOWPVT (2000) in a single testing session to a sample of 20 respondents age 3 to 17 years, the scores were correlated at .95.

Validity:

1. Content validity: developers used most of the items from the EOWPVT (2000) and added 28 items. Based on expert opinion and recommendations, developers continued to include mostly nouns to ensure clarity and permit reliable use with older adults. The developers also included some gerunds, verbs, and modifiers. The new items were of varying difficulty in order to prevent floor effects for the youngest children. Developers used Classical Test Theory (CTT) and Item Response Theory (IRT) during the item selection process and eliminated four items based on analyses.
2. Concurrent validity: EOWPVT-4 scores are distinguishable by age and disability status.

Developers presented data through age 80 showing that EOWPVT raw scores increase with age for 2 through 70 years. T-tests compared the standard scores of individuals identified with one or more types of disabilities to the matched sample means. All five disability groups (Attention Disability, Learning Disability, Reading Disability, Autism, and Language Impairment) scored significantly lower than the matched sample.

Bias Analysis: Developers conducted an analysis of differential item functioning (DIF) using the Mantel-Haenzel procedure for the following subgroups—gender, urban versus rural residence, and race/ethnicity—and then assessed the significance by using the chi-square statistic. No EOWPVT-4 items were found to be biased or were eliminated based on this analysis.

Methods of Scoring: Assessors score items on a pass/fail basis depending on the oral response given by the child. More than one answer may be acceptable. The raw score is the sum of correct responses, and all responses below the basal are counted as correct. Tables in the manual provide instructions for converting raw scores into age-adjusted standard scores, percentile ranks, and age equivalents.

Interpretability: Developers indicate that an individual with formal training in psychometrics should interpret EOWPVT-4 scores. Scoring software made available by the publisher may be

used only with the earlier English and Spanish-Bilingual editions of the EOWPVT, and a score difference analysis is available for children given both assessments.

Training Support: The manual provides thorough instructions for administration and scoring of the assessment.

Adaptations/Special Instructions for Individuals with Disabilities: No information available.

Report Preparation Support: The scoring software for the EOWPVT (2000) provides a summary report based on test results.

References:

Brownell, Rick. *EOWPVT-ROWPVT Scoring Software*. Novato, CA: Academic Therapy Publications, 2000.

Brownell, Rick. *Expressive One-Word Picture Vocabulary Test Manual: Third Edition*. Novato, CA: Academic Therapy Publications, 2000.

Brownell, Rick. *Spanish-Bilingual Edition: Expressive One-Word Picture Vocabulary Test Manual*. Novato, CA: Academic Therapy Publications, 2001.

Martin, Nancy, and Rick Brownell. *Expressive One-Word Picture Vocabulary Test-4 Manual*. Novato, CA: Academic Therapy Publications, 2010.

EYBERG CHILD BEHAVIOR INVENTORY (ECBI) AND SUTTER-EYBERG STUDENT BEHAVIOR INVENTORY-REVISED (SESBI-R), 1999

Authors:

Sheila Eyberg and Donna Pincus

Publisher:

Psychological Assessment Resources
(800) 331-TEST
www.parinc.com

Initial Material Cost:

Contact publisher for cost information

Representativeness of Norming Sample:

The ECBI is standardized on socio-economically diverse Caucasian families and families of other ethnicities with an equal representation of children between the ages of 2 and 16 years. The SESBI-R was originally standardized on lower-middle socioeconomic status (SES) preschoolers. Caution is given regarding the interpretation of cutoff scores in different geographical regions.

Languages:

English¹

¹ Unofficially translated (by universities and other organizations) into Welsh, Australian, Spanish, German, Norwegian, French, and Chinese.

Type of Assessment:

Direct child assessment and parent report

Age Range and Administration Interval:

2 to 16 years

Personnel, Training, Administration, and Scoring Requirements:

Tests can be administered and scored by individuals who do not have clinical training. Completion of the forms requires at least a 6th grade reading level. Training requires familiarizing oneself with the manual and questionnaires. The ECBI and SESBI-R require 10 minutes each to complete and under 5 minutes each to score. Interpreting the scores requires graduate training in psychology, counseling, or a closely related field.

Summary:

Initial Material Cost: not available

Reliability: 3 (.65 or higher)

Validity: 2 (under .5 for concurrent validity and under .4 for predictive validity)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (self-administered and scored by someone with basic clerical skills, but interpreted by a highly trained individual)

Description: The ECBI and SESBI-R are rating scales that assess the severity of conduct problems in children ages 2 through 16 years as well as the extent to which parents and teachers find the behaviors troublesome. The ECBI, which consists of 36 items, is completed by parents and assesses the frequency of disruptive behaviors occurring in the home. The SESBI-R, which consists of 38 items, is completed by teachers and is useful in the assessment of disruptive behaviors in the school setting. Each test provides an Intensity Raw Score and a Problem Raw Score.

Uses of Information: Taken together, the ECBI and SESBI-R have multiple applications, including use as (1) screening measures in the clinical identification of children for the diagnosis and treatment of externalizing behavior problems, (2) screening measures in the identification of “high-risk” children for delinquency prevention programs, and (3) outcome measures in evaluations of treatment programs for conduct-disordered children.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): for the ECBI, .95 for the Intensity scale and .93 for the Problem scale. For the SESBI-R, .98 for the Intensity scale and .96 for the Problem Scale. The Intensity scale had a mean item-to-total correlation of .76, and the Problem scale had a mean item-to-total correlation of .65. (2) Test-retest reliability: for the ECBI, .75 to .86 for the Intensity scale and .75 to .88 for the Problem scale. For the SESBI-R, .87 for the Intensity scale and .93 for the Problem scale. (3) Inter-rater reliability: for the ECBI, .86 for the Intensity scale and .79 for the Problem scale. For the SESBI-R, inter-rater reliability for the Intensity scale was in the .85 to .86 range for the Intensity scale and was .84 to .87 for the Problem scale.

Validity: (1) Concurrent validity: for the ECBI, Problem and Intensity scores were significantly correlated with both the Externalizing scale (.67 and .75, respectively) and the Internalizing scale (.48 and .41, respectively) of the Child Behavior Checklist. Also, the ECBI was found to correlate significantly with the Parenting Stress Index (PSI): the ECBI Problem and Intensity scores were significantly correlated with the PSI Child Domain scores (.45 and .45, respectively). The scores obtained on the SESBI, SESBI-R, and the Revised Edition of the School Observation Scale (REDSOCS) in the regular classroom were significantly related to off-task and inappropriate behavior categories, but not to noncompliance. (2) Predictive validity: For the SESBI-R, correlations between scores and both the child's number of school suspensions and the

number of referrals to the school principal for conduct problems were .26 to .39 one year later and .21 to .36 two years later.

Method of Scoring: Each behavior is rated on two scales: a 7-point Intensity scale assesses how often the behaviors currently occur in the home or school setting (1 means "never," 4 means "sometimes," and 7 means "always") and a Problem scale (Yes/No) identifies whether the child's behavior is problematic for the parent or teacher. Scores are computed by summing the Intensity scale scores on each page for pages 1 and 2. Similarly, the number of "Yes" responses are summed to come up with a separate Problem scale total for pages 1 and 2. To obtain the Intensity Raw score, the Intensity scale scores from pages 1 and 2 are summed. To obtain the Problem raw score, the Problem scale scores from pages 1 and 2 are summed.

Interpretability: Both the ECBI and SESBI-R are continuous in that higher scores on the scale indicate a greater level of conduct-disordered behavior and a greater impact on the parent or teacher. Comparison with normative data in Chapter 3 of the manual allows for more specific examination of the significance of scores, especially in cases where the cutoff score of either scale has been exceeded. T-score conversions for the raw scores are provided in the appendices of the manual (Appendices C, D, E, and F). T-score conversions for either of the scales that are greater than or equal to 60 are clinically significant. T-Scores below 60 are within the normal range. Caution is issued when interpreting SESBI-R

cutoff scores because the author's analyses with the SESBI-R suggest that for the kinds of behaviors assessed, there is wide variability both within and between geographic locations. Therefore, clinicians should adjust cutoff levels based on local norms as needed.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: Two case studies are presented in the examiner's manual.

References:

Eyberg, Sheila, and Donna Pincus. The ECBI & SESBI-R: Eyberg Child Behavior and Sutter-Eyberg Student Behavior Inventory-Revised: Professional Manual. Odessa: Psychological Assessment Resources, 1999.

FUNCTIONAL EMOTIONAL ASSESSMENT SCALE (FEAS), 2001

Authors:

Stanley Greenspan, Georgia DeGangi and Serena Wieder

Publisher:

The Interdisciplinary Council on Developmental and Learning Disorders
www.icdl.com

Initial Material Cost:

Text Book: \$40 for ICDDL members, \$47 for non-members

Additional protocol booklets: \$8

Representativeness of Norming Sample:

None described.

Languages:

English

Type of Assessment:

Direct observation and possible direct child assessment

Age Range and Administration Interval:

7 months to 4 years (research version); Six versions: 7-9

months; 10-12 months; 13-18 months; 19-24 months; 25-35 months; 3-4 years

Personnel, Training, Administration, and Scoring

Requirements: Takes 15-20 minutes to administer. The examiner should be trained and experienced. The authors recommend videotaping the caregiver-child play interaction session. They advise that live scoring should not be attempted without first observing at least 10 videotapes with at least an 80 percent reliability in scoring live and videotaped observations.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (.65 or higher for inter-rater reliability; no other reliability provided)

Validity: 2 (< .5 for concurrent)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Functional Emotional Assessment Scale (FEAS) provides a framework for observing and assessing a child's emotional and social functioning in the context of the relationship with his or her caregiver as well as the caregiver's capacity to support the child's emotional development. The FEAS assesses the child on six levels of social and emotional development: (1) regulation and interest in the world, (2) forming relationships (attachment), (3) intentional two-way communications, (4) development of a complex sense of self, (5) representational capacity and elaboration of symbolic thinking, and (6) emotional thinking or development and expression of thematic play.

There are two versions of the FEAS, a clinical version and a research version. The research FEAS, which evolved from the clinical FEAS, has cutoff scores to assist in interpreting the results and has been used to test for the scale's validity and reliability. Each of these has versions that are designed for different age groups. In both versions, the caregiver (parent) is asked to play with his or her child as he/she might at home for 15 minutes with 3 different types of developmentally appropriate toys: symbolic toys, tactile toys, and toys involving large movement activities. The examiner may also want to engage the child in play to attempt to elicit behaviors not observed during the caregiver-child play

interaction. Because considerable experience is needed to score the FEAS reliably in live observation sessions, the authors recommend that these unstructured play observations be videotaped and scored later. The scale should be used in conjunction with other instruments as part of an overall assessment.

Uses of Information: The FEAS is intended to help clinicians identify critical areas deserving of further clinical inquiry. It can be used descriptively to profile children's emotional, social, and related developmental capacities. It can also be used to diagnosis or screen for problems in children who are experiencing regulatory disorders, but not to formally diagnose specific disorders.

Reliability: (1) Inter-rater reliability (Cronbach's alpha): The alpha coefficients between pairs of observers viewing between 15 and 46 videotaped caregiver-child interactions ranged from .90 to .92 for the caregiver scale and .90 to .98 for the child scale. The alphas between a pair of observers viewing 15 interactions, one coding the interactions live and the other a videotape of the interactions, were .83 for the caregiver scale and .89 for the child scale.

Validity: Four non-nationally representative samples of young children between the ages of 7 and 48 months, except when noted otherwise, were used to test for validity: (1) 197 normal children; (2) 190 children with regulatory disorder; (3) 41 children between the ages of 19 and 48 months with pervasive developmental disorder; and (4) 40 multi-problem children. All

of the samples had a larger proportion of boys, white, and middle-class children. (1) Construct validity:¹ The scores obtained by normative and clinical samples of young children were compared using a discrimination index, t-tests, and analysis of variance. (2) Accuracy of cutoff scores (ranges for the different age groups): False normal errors for the total (child and caregiver) scale ranged from 5 to 28 percent, false delay errors ranged from 26 to 63 percent, specificity (probability correctly identifying a normal child) ranged from 37 to 74 percent, and sensitivity (probability of correctly identifying a delayed child) ranged from 74 to 95 percent. (3) Concurrent: Intercorrelations between the FEAS scores during symbolic and tactile play and two other instruments developed by the authors, the Test of Sensory Functions in Infants and the Test of Attention in Infants, were not significant. The authors interpret this to mean that the FEAS provides unique information.

Method of Scoring: The clinical FEAS may be left unscored and used to provide a descriptive profile of the young child's developmental capacities or to help systematize clinical thinking. The scale can also be used to rank each item as follows: capacity not present (0), capacity fleetingly present (1), capacity intermittently present (2), capacity present most of the time (3), capacity present all of the time in all circumstances (4), or no opportunity to observe capacity (not applicable). The ratings can then be

¹ Note that although the authors consider this information to reflect construct validity, the relationships described are consistent with the way concurrent validity is used throughout this resource guide.

summed for each functioning area and divided by the functioning area's maximum possible score to obtain a percentage. However, only some of the developmental functioning areas can be described quantitatively; the others should be described qualitatively. The research FEAS rates both the caregiver and the child on their mastery of the skill as follows: behavior is not seen or is observed only briefly (skill not mastered) (0), behavior is present some of the time or observed several times (skill partially mastered) (1), and behavior is consistently observed or observed many times (skill mastered) (2). The ratings can be summed to obtain category and subtest scores for the caregiver and category and subtest scores for the child, as well as a combined caregiver total score and a combined child total score.

Interpretability: A developmental growth chart can be used to help assess the child's functional developmental accomplishments over time based on information collected from the clinical FEAS. The research FEAS has cut-off

scores that can be used to determine if parent-child interaction patterns are normal, at risk, or deficient. However, in interpreting both the clinical and research FEAS, the authors strongly recommend that the FEAS not be used alone, but as part of a comprehensive assessment of the caregiver-child relationship.

Training Support: The author offers a training course for the FEAS. Information can be acquired by calling 301-320-6360 or by visiting <http://www.icdl.com/>.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

DeGangi, G. *Pediatric Disorders of Regulation in Affect and Behavior. A Therapist's Guide to Assessment and Treatment.* San Diego, CA: Academic Press, 2000.

HAWAII EARLY LEARNING PROFILE (HELP), 1997

Authors:

Stephanie Parks

Publisher:

VORT Corporation

(650) 322-8282

www.vort.com

Initial Material Cost:

Inside HELP: Administration and Reference Manual, \$50

HELP Family-Centered Interview (0-3),

\$25 (package of 25)

HELP Strands (0-3), \$3 each

HELP Checklist (0-3), \$3 each

HELP Charts (0-3), \$3 each

HELP Activity Guide (0-3), \$28 each

Representativeness of Norming Sample:

None described.

Languages:

English

Type of Assessment:

Direct child or parent assessment

Age Range and Administration Interval:

0 to 36 months

Personnel, Training, Administration, and Scoring

Requirements: HELP is for use by trained professionals. An initial direct assessment may last from 45 to 90 minutes, but in some cases may be completed in 15 to 20 minutes. Most of the scoring is done during the assessment.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scoring by a trained professional)

Description: HELP is a curriculum-based assessment consisting of 685 developmental skills and behaviors covering six traditional child developmental domains: cognitive, language, gross motor, fine motor, social-emotional, and self-help. The developmental skills are organized by skill domains and, within skill domains, by “conceptual strands.” Within each strand, the skills are sequentially ordered by age. For the assessment, the examiner selects several developmental skills above and below the child’s approximate developmental age from each developmental domain and prepares 5 to 10 play or daily activities to elicit several skills concurrently. In addition to working with the child using the prepared activities, the examiner

also observes the child’s environment and interviews the parents for information.

Uses of Information: HELP can be used to identify needs, track growth and development, and develop plans for meeting objectives. It can be used to identify a child’s developmental skills and behaviors along multiple lines of development, the child’s developmental skill and behavior strengths and needs, physical and social environment factors affecting development, and the way development in one area may be affecting development in other areas.

Reliability: None reported.

Validity: Face and content validity tests were used to select and group the skills. The skills were

selected from growth-and-development scales and standardized tests. An interdisciplinary team of pediatric therapists grouped the skills into strands and sequentially ordered them by age. No tests of concurrent or predictive validity reported.

Method of Scoring: Behaviors are scored as “present,” “not present,” “emerging,” “atypical/dysfunctional,” or “not applicable.” HELP Strands or HELP Checklist can be used to record outcomes and track progress. Because it divides each domain into strands, HELP Strands is recommended for children with disabilities or more uneven development within a domain.

Interpretability: Although the manual does not provide any exact rules or formulas for determining approximate developmental levels, it does provide general “rule of thumb” guidelines. The manual provides an explanation for a delay or atypical behavior for each skill area, along with appropriate interventions.

Training Support: A 20-minute training video, *Using HELP Effectively*, can be purchased

from VORT for \$14.95. It is recommended that this video be used in conjunction with the *Inside HELP*. The instruction chapter of the *Inside HELP* manual can be found in the training section on the VORT web page.

Adaptations/Special Instructions for Individuals with Disabilities: *Inside Help* has examples of adjustments to accommodate specific disabilities and special needs when assessing skills and behaviors.

Report Preparation Support: HELP Strands and HELP Checklist can be used to communicate the child’s progress to parents.

References:

Parks, Stephanie. *Inside HELP: Administration and Reference Manual*, 1997 Revision. Palo Alto, CA: VORT Corporation, 1999.

VORT Corporation. *Using HELP Effectively*. Palo Alto, CA: VORT Corporation, 1994

HIGH/SCOPE CHILD OBSERVATION RECORD (COR) FOR AGES 2 1/2-6, 1999

Authors:

High/Scope Education Research Foundation

Publisher:

High/Scope Press
(313) 485-2000
press@highscope.org

Initial Material Cost:

COR Kit: \$125 (includes a Manual, 25 Assessment Booklets, 4 sets of Anecdotal Notecards, 50 Parent Report)

Representativeness of Norming Sample:

No norming sample described.

Languages:

English and Spanish

Type of Assessment:

Observation

Age Range and Administration Interval:

2 1/2 to 6 years; assessment is intended to be a full-year

assessment usually done two to three times throughout the school year.

Personnel, Training, Administration, and Scoring

Requirements: Training in COR for teachers and teaching assistants is recommended. For administering, the manual recommends focusing on a few children each day or two and writing notes on those children specifically. If done by hand, scoring takes approximately one hour. If done on computer, scoring is less than five minutes.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 2 (majority of correlations are < .50)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (administered and scored by someone with basic skills)

Description: With the COR, a trained teacher or observer assesses each child's behavior and activities in six categories of development: (1) initiative, (2) social relations, (3) creative representation, (4) music and movement, (5) language and literacy, and (6) logic and mathematics. Over several months, the teacher writes brief, anecdotal notes describing examples of children's behavior in these six categories. The teacher then uses these notes to rate the child's behavior on 30 five-level COR items within these categories. The COR can be administered at various points throughout the year to measure change over time, or at a single point in time to measure the current developmental level of a child.

Uses of Information: The COR can be used to assess the educational progress of individual children or a group of children as a whole and the program's curriculum's contribution to children's development. It can also be used to develop program plans that focus on specific areas of child development based on the outcomes of the assessment at the individual level. The results can also be shared with the next year's program staff as well as parents.

Reliability: (1) Internal consistency (Cronbach's alphas): ranged from .80 to .93 for teachers and .72 to .91 for assistant teachers. (2) Inter-observer reliability (Pearson correlation coefficients): ranged from .61 to .72.

Validity: (1) Concurrent validity: correlations of the COR development categories with similar categories on the McCarthy Scales of Children's Abilities ranged from .27 to .53. The correlations between the COR and all of the McCarthy Scale categories ranged from .27 to .66.

Method of Scoring: Items are scored by the six categories. Using the anecdotal notes, the highest level of behavior that is characteristic of the child is checked. The levels go from one through five, with five being the highest score possible for each item. If scoring by hand, there is a formula to use; if scoring on the computer, the computer does it for you.

Interpretability: The COR is meant to follow a child's (or children's) development over time, for instance, over the school year.

Training Support: The COR manuals have written support with examples. However, High/Scope recommends that teachers participate in a two- or three-day workshop on the use of COR offered throughout the country. The training covers how to recognize developmentally significant behavior and describe it in anecdotal

notes, how to select the item and item level that each anecdotal note represents, and how to report these results to parents and program officials. The training is \$190 per person for two days, and \$95 per person for the additional computer-training day. Also, those individuals who go through training are given a practice CD, which is used as a follow-up/refresher once training has been completed. Training information and a schedule is available on the High/Scope web page, <http://www.highscope.org/>.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: Parent Report Forms are included in the package and are used as the basis for discussion at parent conferences.

References:

High/Scope Educational Research Foundation. High/Scope Child Observation Record (COR) For Ages 2 1/2-6. Ypsilanti, MI: High/Scope Press, 1992.

HIGH/SCOPE CHILD OBSERVATION RECORD FOR INFANTS AND TODDLERS (COR-IT), 2002

Authors:

High/Scope Educational Research Foundation

Publisher:

High/Scope Press
(800) 407-7377
www.highscope.org

Initial Material Cost:

COR-IT Kit (includes 3 Observation Items Booklets, User Guide, 25 Anecdote Forms, 25 Observations About Your Child Forms, 25 Parent Guides, 25 Child Information and Development Summary Forms, 5 Group Summary Forms, and posters): \$150

Representativeness of Norming Sample:

None described.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

6 weeks to 3 years. Administer 2 to 3 times during a program year.

Personnel, Training, Administration, and Scoring

Requirements: No special degrees are required to conduct the observation, however, it is recommended that the observer receive a two-day training by High/Scope and be in regular contact with the children being observed. The observer should do the observation over a period of several weeks or months and spend 30 to 60 minutes writing rough notes and transforming them into formal notes.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 and higher for both internal and inter-rater reliabilities)

Validity: 3 (.5 and higher for concurrent validity)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a trained individual)

Description: The High/Scope Child Observation Record for Infants and Toddlers (COR-IT) enables the user to conduct an ongoing, comprehensive, systematic assessment in programs serving children from the ages of 6 weeks to 3 years. COR-IT's 28 items are divided into 6 categories: (1) sense of self, (2) social relations, (3) creative representation, (4) movement, (5) communication and language, and (6) exploration and early logic. Over weeks or months, the caregiver records brief, anecdotal notes describing examples of children's behavior in these six categories. At the end of the observation period, which usually takes place over a period of weeks or months, the caregiver uses these notes to complete a development summary

form. The COR-IT has a computer software version that allows individuals to use their computers to record and store observations and to generate COR scores and reports based on this information.

Uses of Information: COR-IT provides an accurate assessment of a child's development and abilities, which can be used to help programs plan activities for the child, to monitor the effects of these activities and the progress of the child, to assess the effectiveness of the program and its curriculum, and to provide a framework for communicating with parents about a child's developmental needs.

Reliability: (1) Internal consistency reliability (Cronbach alpha): The alpha for the entire 28-item scale was .99, and the alphas for the six categories were .92 or .93. The alphas for the 28-item scale for each age category were .94 for infants under 1 year, .95 for children 1 to 2 years, and .78 for children 2 to 3 years. (2) Inter-rater reliability: The Pearson correlations between two groups of observers were .93 for the overall scale and ranged from .83 to .91 for the six categories.

Validity: (1) Concurrent: The correlations between COR-IT and the Bayley Mental and Motor Age Scores were .87 and .91, respectively. The correlations ranged from .88 to .92 between the six COR-IT categories and the motor age score and .83 to .90 between the categories and the mental age score. Since both the COR-IT and the Bayley scores are strongly influenced by the child's age, the authors also tested for validity with the effects of age statistically removed from the correlations. When they did this, they obtained correlations between the COR-IT scale and the mental and motor scores of .26 and .36 respectively.

Method of Scoring: The Observation Item manual provides 5 examples of typical behavior for each of the 28 items. For each item, the caregiver compares those examples with his/her notes to rank the child's typical behavior of the item on a five-point scale from simple (1) to more complex (5) in the development summary form and also enters the highest level of behavior the child achieved. The form provides instructions to compute the average and composite scores. If the

program requires a group summary, the caregiver completes the group summary form using the information on the children's development summary form and follows the instructions on the form to compute the average and group growth scores.

Interpretability: The results from the observations are intended to be used by both the care-giver and the program administrators. No descriptions are provided on how to interpret the results.

Training Support: High/Scope recommends the caregiver attend a two-day workshop on the use of COR-IT, which is offered throughout the country. There is also a one-day training on the computer software. Information on the training schedule and topics are available on the High/Scope Website or by emailing: training@highscope.org.

Adaptations/Special Instructions for Individuals with Disabilities: Adaptations and special instructions are unnecessary as long as the child's level of functioning is between ages 6 and 36 months. However, High/Scope cautions that the reliability and the validity of the instrument with special needs populations have not been established.

Report Preparation Support: An Observation About Your Child form can be used to prepare a report for parents. The COR-IT computer software allows for presentation of ratings in both graphic and narrative form.

References:

High/Scope Educational Research
Foundation. High/Scope Child Observation
Record for Infants and Toddlers User Guide.
Ypsilanti, MI: High/Scope Press, 2002.

HUMANICS NATIONAL INFANT-TODDLER ASSESSMENT, 1994

Authors:

Jane A. Caballero and Derek Whordley

Publisher:

Humanics Psychological Test Corp.

(800) 874-8844

(404) 874-1976 (fax)

www.humanicslearning.com

info@humanicslearning.com

Initial Material Cost:

Manual: \$19

Forms (25): \$35

Representativeness of Norming Sample:

None. The authors recommend that each facility develop local norms and provides the formula to derive percentage rank.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

0 to 3 years; four-month intervals are recommended between assessments, but the examiner can choose to deviate from this.

Personnel, Training, Administration, and Scoring

Requirements: Can be administered by teachers or parents. Training to enhance observational skills is desirable, and some knowledge of child development concepts is required to understand the importance of assessment items.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (administered and scored by someone with basic clerical skills)

Description: The Humanics National Child Assessment Form (HNCAF) is a checklist of 90 skills and behaviors exhibited by children during the first three years of life. The skills/behaviors are grouped into four broad categories—social-emotional, language, cognitive, and gross and fine motor development. The assessment lists 18 skill/behavior items in each of these five areas, listed in the order they are likely to emerge. The assessment can be administered either through informal observations or through activities with the child.

Uses of Information: HNCAF is designed for use by caregivers to identify the skills and behavior a child has, to screen for developmental or physical impairments, to plan learning

experiences that facilitate further growth, and to monitor the child's progress. HNCAF can also be used for parent training.

Reliability: Not discussed.

Validity: Not discussed.

Method of Scoring: The child's responses are scored by checking one of two boxes (occurs consistently or occurs occasionally) or, if the behavior is not present, by leaving the boxes unchecked.

Interpretability: The manual discusses how the assessment results can be used to develop an individual profile and educational plan for the

child and has an example of how this might be done.

Training Support: Not mentioned in the manual. The authors suggest reading Betty Rowen's book, *The Children We See* (Holt, Rinehart and Winston, Inc., 1973) for techniques for observing children's behavior.

Adaptations/Special Instructions for Individuals with Disabilities: The manual describes areas of disabilities and indicators of these disabilities that may be observed while administering the assessment.

Report Preparation Support: The manual provides an example of a completed report.

References:

Caballero, Jane and Derek Whordley.
Humanics National Infant-Toddler Assessment Handbook: A User's Guide for the Humanics National Child Assessment For Ages 0-3.
Atlanta, GA: Humanics Psychological Test, Corp., 1994.

INFANT-TODDLER DEVELOPMENTAL ASSESSMENT (IDA), 1995

Authors:

Sally Provence, Joanna Erikson, Susan Vater, and Saro Palmeri

Publisher:

Riverside Publishing
(800) 323-9540
www.riverpub.com

Initial Material Cost:

IDA Complete Kit: \$502 (includes 25 Parent Reporting Forms, Health Record Guides, and Record Forms, a Foundations and Study Guide, the Administration Manual, readings, and IDA Manipulative Kit in Carrying Case) Without Manipulative Kit and Carrying Case: \$278

Representativeness of Norming Sample:

The research sample of 100 children between birth and 3 years old is not nationally representative.

Languages:

English¹

¹ The parent report is also available in Spanish.

Type of Assessment:

Parent report and observation

Age Range and Administration Interval:

Birth to 42 months

Personnel, Training, Administration, and Scoring Requirements:

A multidisciplinary team or a very well-trained clinician can administer the assessment. Training tapes are available. Administration and scoring times vary.

Summary:

Initial Material Cost: 3 (> \$200)

Reliability: 3 (.65 or higher)

Validity: 3 (based on percent scoring agreement with other instruments)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Infant-Toddler Developmental Assessment (IDA) is designed to improve early identification of children birth to 3 years of age who are developmentally at risk. There are six IDA phases that are designed to be conducted by a team of two or more professionals: (1) Referral & Pre-interview Data Gathering, (2) Initial Parent Interview, (3) Health Review, (4) Developmental Observation and Assessment, (5) Integration and Synthesis, and (6) Share Findings, Completion, and Report. Each phase develops from the preceding one and is completed only after team discussion and review. The Developmental Observation and Assessment

Phase (Phase Four) uses the Provence Birth-to-Three Developmental Profile, which provides a descriptive summary of a child's developmental competencies. The entire assessment uses observation and parent reports of the child's development along eight developmental domains: Gross Motor, Fine Motor, Relationship to Inanimate Objects (Cognitive), Language/Communication, Self-Help, Relationship to Persons, Emotions and Feeling States (affects), and Coping. The Provence Profile is to be used within the context of the full IDA rather than as an isolated test. Five forms are used to gather and record information: Parent Report,

Request for Health Information, Family Recording Guide, Health Recording Guide, and IDA Record.

Uses of Information: The IDA helps determine the need for monitoring, consultation, intervention, or other services for the child and family and may be used to develop an Individualized Family Service Plan (IFSP).

Reliability: (1) Internal consistency reliability: alpha coefficients for the Provence domain scores range from .90 to .96 for ages 1 to 18 months and .77 to .96 for ages 19 to 36 months. (2) Test-retest reliability: no information available. (3) Inter-rater reliability: correlations between raters ranged from .91 to .95 for seven of the eight domains and .81 for the remaining domain (language/communication).

Validity: (1) Concurrent validity: comparisons between the IDA and the Bayley Scales of Infant Development, Hawaii Early Learning Profile, Learning Accomplish Profile, and the Vineland Adaptive Behavior Scales showed that, of the items IDA had in common with those instruments, the percentage agreement on the developmental age ranged from 84 to 100 percent. A comparison of the IDA with the Bayley and Vineland Adaptive Behavior Scales in identifying children needing services found the IDA and Vineland had more similar outcomes. The IDA classified 51.6 percent of the children as needing services compared to 66.0 to 73.6 percent for the Vineland and 13.2 to 22.6 percent for the Bayley. Of 57 children referred to services by IDA practitioners, according to the receiving agencies,

83 percent of the referrals were considered to be appropriate. (2) Predictive validity: no information available.

Method of Scoring: The Provence Protocols behavioral items are marked “present and observed,” “not present or observed,” “reported present and not observed,” “reported not present,” “emerging,” or “refused.” The number of correct responses is used to determine the child’s performance age, which is compared to the child’s chronological age to determine whether to rate the child’s development in the domain as “competent” or “of concern” and, if the latter, the degree of the delay. Using tables in the manual, the “Percentage Delay” can also be computed from the child’s observed performance age and the child’s chronological age, adjusted for prematurity.

After obtaining the Provence Protocol score, the assessment team needs to take into account qualitative aspects of the child’s performance and performance on certain “marker” skills for the child’s age group to decide again whether the child’s development is “competent” or “of concern” and, if the latter, the level of concern.

Interpretability: The manual provides a general guideline on how to interpret the results and what should follow based on the results.

Training Support: Training materials are available to help professionals train others on the administration of the IDA. Materials consist of a Leader’s Guide and three videos. These materials are not meant to be a “self-study” course for individuals trying to learn IDA. The Erikson

Institute also provides training. For more training options, visit www.erickson.edu/.

Adaptations/Special Instructions for Individuals with Disabilities: IDA meets the criteria for assessment as required by the regulations for the Individuals with Disabilities Education Act (IDEA).

Report Preparation Support: The Manual gives instructions on how to share findings and develop a plan with parents in Phase Six, “Share Findings, Completion, and Report.”

References:

Provence, S., J. Erikson, S. Vater, and S. Palmeri. Infant-Toddler Developmental Assessment (IDA) Administration Manual. Itaska, IL: Riverside Publishing, 1995.

<http://www.erickson.edu/>

INFANT-TODDLER SOCIAL AND EMOTIONAL ASSESSMENT (ITSEA) AND BRIEF INFANT-TODDLER SOCIAL AND EMOTIONAL ASSESSMENT (BITSEA), 2006 (2011 Update)

Authors:

Alice S. Carter and Margaret J. Briggs-Gowan

Publisher:

Pearson Assessments
(800) 627-7271

<http://www.pearsonassessments.com>

ITSEA Instrument:

<http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8007-387>

BITSEA Instrument:

<http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8007-352>

Initial Material Cost:

ITSEA—Complete Kit with Scoring Assistant includes ITSEA Manual, 25 Parent Forms, 25 Childcare Provider Forms, Scoring Assistant: \$223.35

BITSEA—Kit includes BITSEA Manual, 25 Parent Forms, 25 Childcare Provider Forms: \$107

Representativeness of Norming Sample:

The ITSEA norming sample was a national, stratified sample of 600 toddlers age 12 to 36 months. The sample consisted of four age groups, with 150 toddlers per group. Based on the 2002 U.S. Census, each age group was stratified along the following variables: gender, region, ethnicity, and parent education. The toddlers were recruited from Women, Infants, and Toddlers (WIC) program sites, doctor's offices, apartment complexes, and community centers. The BITSEA was normed by using the same standardization sample. The French and Chinese speaking samples are described below (see Other Languages).

Languages:

English, French, Spanish, and Chinese

Type of Assessment: Parent or toddler care provider report (self-administration) or direct parent assessment (structured interview)

Age Range and Administration Interval:

12 through 35 months

Personnel, Training, Administration, and Scoring Requirements:

Both the ITSEA and the BITSEA may be self-administered by parents or child care providers or administered and scored by teachers, clinicians, or professionals experienced with standardized tests and developmental and mental health training specific to working with toddlers and families. Parents may administer the assessments under the guidance of a qualified professional. A professional with training in standardized assessment interprets results, although parents and others not highly trained may interpret the assessment under supervision.

ITSEA administration time is 20 to 30 minutes as a questionnaire and 35 to 45 minutes as an interview. The BITSEA Parent Form takes 5 to 7 minutes as a questionnaire and 7 to 10 minutes as an interview.

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 3 (mostly .65 or higher)

Validity: 2 (mostly under .5 for concurrent); 1 (none described for predictive)

Norming Sample Characteristics: 3 (normed within past 15 years and nationally representative)

Ease of Administration and Scoring: 3 (self-administered or administered by a child care provider and scored by a trained individual)

Description: The Infant-Toddler Social and Emotional Assessment (ITSEA) detects social-emotional and behavior problems and delays in the acquisition of competencies in toddlers 12 through 35 months of age. Parents and child care providers observe toddlers in natural

environments and answer questions on the Parent Form or the Childcare Provider Form.¹ The

¹ The caregiver may complete a Childcare Provider Form as a self-conducted questionnaire to supplement the Parent Form. The Childcare Provider form is not normed and should be used only as a basis of comparison for loosely interpreting inter-rater agreement.

Parent Form may be administered as either a self-conducted questionnaire or a structured interview with questions read verbatim to parents. The 168-item ITSEA measures four behavioral domains² that encompass 17 subscales: (1) Externalizing using activity/impulsivity, aggression/defiance, and peer aggression subscales (24 items); (2) Internalizing using depression/withdrawal, general anxiety, separation distress, and inhibition-to-novelty subscales (32 items); (3) Dysregulation using sleep, negative emotionality, eating, and sensory sensitivity subscales (34 items); and (4) Competence using compliance, attention, imitation/play, mastery motivation, empathy, and prosocial peer relations subscales (37 items). Taken together, Externalizing, Internalizing, and Dysregulation are termed “Problem domains.” The ITSEA also includes three item clusters—Maladaptive (13 items), Atypical (8 items), and Social Relatedness (10 items)—to identify more serious problems. Ten Individual Items of Clinical Significance do not belong to any subscale. The Maladaptive item cluster identifies behaviors related to trauma, tic disorders, trichotillomania, elimination problems, and pica. The Atypical item cluster identifies behaviors linked with autism spectrum disorders. The Social Relatedness item cluster categorizes deficits in competencies often associated with toddlers with autism spectrum disorders. In contrast to the ITSEA Parent Form, the Childcare Provider Form has one fewer item in the Dysregulation domain and does not include the

² All domains and item clusters except for Competence screen for potential problems.

extra section of non-subscale items of clinical significance. The previous version of the ITSEA (1998) was updated with a more current norming sample, and Parent Form item sequencing was modified before collection of the norming data.

The Brief Infant-Toddler Social and Emotional Assessment (BITSEA) is a short 42-item version to screen for whether a toddler requires a more in-depth social-emotional evaluation. The age range is the same as the ITSEA: 12 months to 35 months, 30 days. The BITSEA consists of a Parent Form that may be completed as a self-administered questionnaire or read aloud verbatim as an interview with the parent. The BITSEA addresses the domains of Internalizing (8 items), Externalizing (6 items), and Dysregulation (8 items) as well as less common behaviors that may indicate autism spectrum disorders (17 items) or other psychopathologies (14 items). Eleven items also assess the acquisition of social-emotional competencies. Several items measure more than one domain.

Other Languages: The ITSEA and BITSEA are available in Spanish for use with U.S. populations. The norming sample, reliability, validity, and English language equivalence for the Spanish versions are unavailable. The ITSEA is available in French and Chinese for use with the respective country populations. The reliability and validity of the French version were studied with a sample of French parents of 250 toddlers recruited from infant wellness clinics and childcare centers (Bracha et al. 2004). Confirmatory factor analyses

validated the use of the ITSEA to detect toddler delays or deficits and assess social-emotional competency. Most subscales and all domains showed adequate intrascale reliability, but some gender and age effects on mean scores differed from those reported for the U.S. sample. Age effects were observed for Eating Problems, Sleep, and Inhibition to Novelty in the French sample, but not in the U.S. sample. For the French sample only, girls were rated higher than boys on both Negative Emotionality and Aggression/Defiance. The reliability and validity of the Chinese version of the ITSEA Parent Form were examined with a sample of 5,323 toddlers from 14 Chinese cities (Jianduan et al. 2009). Test-retest reliability ranged from .78 to .89. Internal consistency was reported with alpha coefficients ranging from .79 to .88. Concurrent validity was also demonstrated with correlations between the Chinese ITSEA and two other Chinese assessments measuring similar constructs. Confirmatory factor analysis showed that the Chinese ITSEA could be used for reliably assessing the social-emotional problems and competencies of urban Chinese toddlers.

Uses of Information: The ITSEA identifies toddlers who may have abnormal behaviors, psychopathology, or delayed competencies. To establish a toddler's eligibility for services, the ITSEA should be supplemented by direct observation, additional assessment, and/or discussion with parents. The BITSEA is a shorter version of the ITSEA that determines whether a more in-depth social-emotional assessment is warranted.

Reliability:

1. Internal consistency: Cronbach's alphas were reported by age group, gender, and domain for the ITSEA. Externalizing alpha coefficients were .88 for girls and .86 for boys and ranged from .66 to .79 across subscales, age groups, and gender. Internalizing coefficients were .85 for both girls and boys and ranged from .52 to .73 across subscales, age groups, and gender. Dysregulation coefficients were .86 for both girls and boys and ranged from .62 to .83 across subscales, age groups, and gender. Competence coefficients were .89 for girls and .90 for boys and ranged from .56 to .79 across subscales, age groups, and gender. More than half of the coefficients for the subscale exceeded .70.
2. Test-retest reliability: ITSEA estimates were based on a sample of 84 parents assessed or interviewed by using the Parent Form within a two- to eight-day interval (mean of six days) between administrations. The ITSEA domain and subscale coefficients were .90 for Externalizing (.80 to .91 for the subscales), .85 for Internalizing (.78 to .83 for the subscales), .91 for Dysregulation (.81 to .92 for the subscales), and .76 for Competence (.75 to .85 for the subscales). The coefficients for the item clusters were .77 for Maladaptive, .64 for Social Relatedness, and .83 for Atypical. BITSEA estimates were based on the same sample of 84 toddlers over the same interval. Pearson's correlation coefficients across toddler gender groups and domain scores

were all above .80. Coefficients for the total Problem domain score (that is, an aggregate score for Internalizing, Externalizing, and Disregulation) and total Competence score were .92 and .82, respectively.

3. Inter-rater reliability for the ITSEA: Intraclass Correlation Coefficients (ICC) to show agreement between 94 pairs of parents were .77 for Externalizing (.68 to .77 for the subscales), .72 for Internalizing (.51 to .73 for the subscales), .77 for Dysregulation (.61 to .81 for the subscales), and .79 for Competence (.51 to .83 for the subscales). More than half of the subscale scores exceeded .70. The ICCs for the item clusters were .70 for Maladaptive, .44 for Social Relatedness, and .59 for Atypical. The BITSEA inter-rater sample of 94 pairs of parents showed ICCs for problem domain scores and Competence scores at .74 and .63, respectively.

Validity:

1. Content validity: For the original ITSEA (1998), the authors conducted an extensive literature search on social-emotional development and behaviors and explored existing measures for older toddlers. They developed a list of 200 items that were reviewed by toddler development professionals with expertise in social-emotional development. The experts suggested additional items to improve the instrument's content coverage. The authors had previously conducted a confirmatory factor analysis on the domains and subscales

of the measure (Briggs-Gowan and Carter 2005). That analysis validated the structure of the instrument as hypothesized by the authors. The revised items underwent pilot testing and another revision to make them more understandable and user-friendly. The authors conducted three field tests to assess the preliminary reliability and validity of the measure and further refine the items. An expert panel also identified which items from the ITSEA should be included in the BITSEA (Geisinger and Murphy 2007).

2. Concurrent validity:³ the ITSEA Parent Form was compared to the BITSEA as well as to several similar social-emotional assessments, including the Toddler Behavior Checklist 1.5-5 (CBCL 1.5-5), the Ages and Stages Social-Emotional Questionnaires: Social Emotional (ASQ:SE), the Adaptive Behavior Assessment System: Second Edition (ABAS-II), and the Bayley Scales of Infant and Toddler Development-Third Edition (Bayley-III).

The ITSEA Problem domain scores that correlated with the BITSEA total Problem score ranged from .57 for Internalizing to .77 for Dysregulation. The ITSEA Parent Form Competence domain correlated with the BITSEA total Competence score at .69 for

³ Concurrent validity toddler samples were distinct for each measure comparison, although each sample was ethnically diverse and half female. The sample sizes for each comparison group are as follows: CBCL, n= 37 (12 to 17 month olds were excluded); ASQ:SE, n= 57; ABAS-II, n= 52 (of which only 6% were 18 to 23 month-olds); and Bayley-III, n= 112.

girls and .77 for boys. The correlation coefficient between the ITSEA Externalizing domain and CBCL Externalizing composite was .41. Internalizing Problem scores from both assessments were correlated at .60. The ITSEA Dysregulation domain correlated with CBCL Internalizing and CBCL Externalizing composites at .44 and .39, respectively. The ASQ:SE total score and scores on the ITSEA Externalizing, Dysregulation, and Internalizing domains were correlated at .69, .53, and .34, respectively. The correlations between the ITSEA Competence domain and the Bayley-III Cognitive Assessment, Language Composite, and Social-Emotional subtest were .32, .33, and .48, respectively. The correlation coefficient between the ITSEA Competence domain and the General Adaptive Behavior composite on the ABAS-II was .52. Authors predicted low correlation coefficients between the ITSEA and other comparison measures based on different constructs. The ITSEA Problem domain scores and the Adaptive Behavior scores on the ABAS-II correlations were .06 for Externalizing, -.31 for Internalizing, and -.13 for Dysregulation. The ASQ:SE total score and the ITSEA Competence domain were correlated at -.43.

Authors compared the BITSEA to the ASQ:SE, CBCL, ABAS-II, and Bayley-III. The BITSEA total Problem score was correlated with the ASQ:SE total score at .55, CBCL Internalizing at .46, CBCL Externalizing at .47, and the CBCL total at .6. The BITSEA

total Competence score correlated with the ABAS-II Social composite, Practical composite, and General Adaptive Composite at .56, .39, and .44, respectively. The BITSEA total Competence and the Bayley-III Cognitive Assessment, Language scale, and Social-Emotional scores had coefficients of .25, .32, and .51, respectively. To examine differences in measures, authors correlated the BITSEA total Problem score with the ABAS-II Practical, Social Skills, and Conceptual composites, which had coefficients of -.31, -.36, and -.36, respectively, and with the Bayley Social-Emotional, Cognitive Assessment, and Language scale, which correlated at -.27, -.28, and -.19. The BITSEA total Competence score correlated with the ASQ:SE total score at -.55 and with the CBCL Internalizing, Externalizing, and total scores at -.30, -.42, and -.38.

Authors of the ITSEA and BITSEA examined differences between toddlers who were “typically developing” and toddlers with autism, delays, or premature birth or who had received a mental health service referral. Competence domain score differences distinguished all toddlers except those born prematurely, although the authors note that scores are adjusted for prematurity. The authors conducted research showing that, based on differences in Social Relatedness and Atypical item cluster scores, the ITSEA significantly differentiates autistic toddlers from those with a developmental delay and those developing typically. The BITSEA

distinguished autistic toddlers based on total Competence and total Problem score differences.

3. Construct validity: intercorrelations between ITSEA subscale and domain scores ranged from .50 to .69 between Externalizing domain and associated subscales scores, from .34 to .42 for the Internalizing domain and subscales, from .32 to .47 for the Dysregulation domain and subscales, and from .62 to .69 for the Competence domain and subscales. Correlations were corrected by removing the subscale score from the raw domain score. Correlations among the Externalizing, Internalizing, and Dysregulation domains ranged from .43 to .53 while correlations between the Competence domain and the other domains ranged from -.15 to -.27.

Bias Analysis: No information available.

Methods of Scoring: The ITSEA and the BITSEA are scored by professionals experienced in standardized testing and trained in developmental and mental health specific to toddlers and families. ITSEA and BITSEA items are rated on a three point scale: 0 = not true/rarely, 1 = somewhat true/sometimes, and 2 = very true/often; caregivers may also select N = “no opportunity” if they did not observe the behavior. Before scoring the ITSEA, scorers code “no opportunity” responses as missing; the responses thus are not reflected in a toddler’s score. If a subscale exceeds a maximum number of missing items, then that subscale is not be scored.

Seven items are recoded or “reverse scored” to ensure that the given problem or competence is appropriately captured. For example, the item *Is able to wait for things he or she wants* on the Negative Emotionality subscale is recoded so that a score of 0 becomes 2, a score of 1 remains 1, and a score of 2 becomes 0. The following scores may be calculated: domain scores, subscale scores, and item cluster scores. Domain scores equal the raw subscale scores divided by the number of subscales scored. Subscale and item cluster scores are calculated as means equal to the sum of item responses (i.e., 0, 1, or 2) divided by the total number of non-missing items for that subscale or item cluster. The manual provides tables to derive T scores for the domains and cut-points for the subscales and item clusters. Subscale scores may also be converted to percentile ranks. A Score Summary is used to record ITSEA scores and convey information to parents about a toddler’s performance (see Report Preparation Support).

To score the BITSEA, the assessor transfers parent responses to a Score Summary Form and sums the Problem domain items and Competence items for total Problem and Competence scores. Unanswered items, including missing and no opportunity responses, do not count toward total Problem or Competence scores. The assessor records cut scores from the manual next to total scores on the Score Summary Form. The manual lists percentile rankings that correspond to a toddler’s total Competence and total Problem scores.

Interpretability: The ITSEA should be interpreted by a professional who has received training in the administration and interpretation of psychometric tests or by parents or other caregivers with the help of a professional. Tables in the manual may be used to convert mean raw subscale scores into percentile rankings by age and gender. Percentile rankings correspond to score ranges for each subscale. Subscale scores falling at or below the 10th percentile are deemed to be in a range “of concern.” The manual provides a table that lists the number of subscales falling into the range “of concern” for the normative sample by sex and age group. It provides a similar table for five clinical groups (autistic disorder, developmentally delayed, language delayed, mental health, premature/low-birth weight). Additional assessment or intervention may be warranted if a toddler receives more subscale scores in the “of concern” range than occur in the normative sample. If rated as occurring sometimes or often, 10 Individual Items of Clinical Significance (that do not pertain to any domain, subscale, or item cluster) warrant followup with the parents to assess the behaviors in more detail. The manual provides three case studies as examples of how to interpret the ITSEA.

For the BITSEA, tables in the manual present percentile rankings that correspond to total raw scores by gender and age. High total Problem scores (at the 75th percentile or higher) and/or low scores (at the 15th percentile or lower) may indicate problems, delays, or deficits. Follow-up

assessments are needed to determine clinical significance.

Training Support: No information available.

Adaptations/Special Instructions for Individuals with Disabilities: If a toddler is unable to perform a particular behavior due to a constraint such as a physical limitation, the caregiver may rate the item as no opportunity instead of assigning 0 to that behavior.

Report Preparation Support: The Score Summary Form for the ITSEA allows the assessor to record scores for each domain and subscale with the corresponding percentile range, item cluster scores (without corresponding percentiles), and comments. The assessor may also use checkboxes to indicate whether a subscale or item cluster score is “of concern.” The manual instructs assessors in how to hand-score the ITSEA by using the Parent Form Hand Scoring Template. Alternatively, the Scoring Assistant CD-ROM software program is available to generate score reports for each subscale, domain, and item cluster. The Score Summary Form for the BITSEA includes Problem domain and total Competence scores as well as percentile rankings and a checkbox where the assessor indicates a possible problem, delay, or deficit.

References:

Baxter, Abigail. “Review of the ITSEA/BITSEA: Infant-Toddler and Brief Infant-Toddler Social and Emotional Assessment.” In *The Seventeenth Mental Measurements Yearbook*, edited by Kurt F. Geisinger, Robert A. Spies, Janet

F. Carlson, and Barbara S. Plake. Lincoln, NE: Buros Institute of Mental Measurements, 2007.

Bracha, Z., F. Perez-Diaz, P. Gerardin, Y. Perriot, F. De La Rocque, M. Flament, M.C. Leroux, P. Reinert, P. Mazet, and A.S. Carter. "A French Adaptation of the Infant-Toddler Social and Emotional Assessment." *Infant Mental Health Journal*, vol. 25, no. 2, 2004, pp. 117-129.

Briggs-Gowan, Margaret J., and Alice S. Carter. *BITSEA Brief Infant-Toddler Social and Emotional Assessment-Examiner's Manual*. San Antonio, TX: Pearson, 2006.

Briggs-Gowan, Margaret J., and Alice S. Carter. "Preliminary Acceptability and Psychometrics of the Infant-Toddler Social and Emotional Assessment (ITSEA): A New Adult-Report Questionnaire." *Infant Mental Health Journal*, vol. 19, no. 4, 1998, pp. 422-445.

Briggs-Gowan, M.J. and A.S. Carter. *Infant-Toddler Social and Emotional Assessment/Brief Infant-Toddler Social and Emotional Assessment*.

San Antonio, TX: Harcourt Assessment, Inc., 2005.

Carter, Alice, and Margaret Briggs-Gowan. *ITSEA Scoring Assistant*TM. San Antonio, TX: Pearson, 2005.

Carter, Alice S., and Margaret J. Briggs-Gowan. *ITSEA Infant-Toddler Social and Emotional Assessment-Examiner's Manual*. San Antonio, TX: Pearson, 2006.

Jianduan, Z., W. Huishan, S. Shuhua, H. Xiaonan, L. Guoyan, L. Guangli, and S. Junxin. "Reliability and Validity of Standardized Chinese Version of Urban Infant-Toddler Social and Emotional Assessment." *Early Human Development*, vol. 85, no. 5, 2009, pp. 331-336.

Konold, Timothy R. "Review of the ITSEA/BITSEA: Infant-Toddler and Brief Infant-Toddler Social and Emotional Assessment." In *The Seventeenth Mental Measurements Yearbook*, edited by Kurt F. Geisinger, Robert Spies, Janet F. Carlson, and Barbara S. Plake. Lincoln, NE: The Buros Institute of Mental Measurements, 2007.

INFANT/TODDLER SYMPTOM CHECKLIST (ITSC), 1995

Authors:

Georgia DeGangi, Susan Poisson, Ruth Sickel, and Andrea Santman Wiener

Publisher:

Therapy Skill Builders, a division of the Psychological Corporation
800-872-1726

Initial Material Cost:

Complete set (Includes manual, 6 sets of 5 score sheets in 25 page pads and vinyl storage portfolio) \$63

Representativeness of Norming Sample: No norming sample.

Languages:

English

Type of Assessment:

Parent report or interview

Age Range and Administration Interval:

7-30 months old

Personnel, Training, Administration, and Scoring Requirements:

Administration time is 10 minutes, and it can be administered by the parent or by a paraprofessional. Special training is not required for administering the instrument; however, an understanding of the domains is critical for an accurate interpretation of findings. Scoring can be done in less than 10 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described)

Validity: 2 (less than .5 for concurrent)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered or administered and scored by someone with basic clerical skills)

Description: The Infant/Toddler Symptom Checklist (ITSC) is designed to screen 7 to 30-month-old infants and toddlers for sensory and regulatory disorders who are behaviorally problematic and show disturbances in sleep, feeding, state control, self-calming, and mood regulation. The checklist focuses on infant behavior in the following domains: (1) self-regulation, (2) attention, (3) sleep, (4) eating or feeding, (5) dressing, bathing, and touch, (6) movement, (7) listening and language, (8) looking and sight, and (9) attachment/emotional functioning. Questions are answered with a “never or sometimes,” “most times,” or “past.” The criterion-referenced ITSC checklist comes in six versions: a single short version for general screening purposes and five age-specific screens for both diagnostic and screening purposes: 7 to 9

months, 10 to 12 months, 13 to 18 months, 19 to 24 months, and 25 to 30 months. The authors recommend using other observation tools when using ITSC as a screening tool and traditional developmental tests when using it as a diagnostic tool.

Uses of Information: The ITSC is used to determine whether a child may have a predisposition toward developing sensory integrative disorders, attention deficits, or emotional, behavioral, or learning difficulties, and whether further diagnosis is required.

Reliability: None described.

Validity: The research sample consisted of 154 normal and 67 regulatory-disordered infants between 7 and 30 months who were primarily

white and middle class. (1) Concurrent validity: Statistical tests were performed on correlations between scores on the ITSC and the Bayley Scales of Infant Development, Mental Scale; the Test of Sensory Functions in Infants (TSFI); and the Test of Attention in Infants (TAI) for a sample of normal infants and a sample of regulatory disorder infants. The results showed that correlations were statistically significant for the regulatory disorder infants and only a few of the correlations with the TSF and TIA subtests were significant for the normal infants, especially among the 7- to 9-month-olds. The authors concluded that the ITSC provided information that is distinct from that obtained by diagnostic measures, particularly for 10- to 30-month olds. The authors tested for construct validity to select the instrument items by performing t-tests on the difference between means obtained from a sample of regulatory disorder infants and a sample of normal infants. They also performed another test by comparing the scores of parents who did not express concern over their infant's development with those of parents who did express concern and found that only one (out of 25) of the scores in the no concern group was above the at-risk cutoff score while all but two (out of 14) in the concern group had scores about the cutoff. The authors then performed diagnostic tests on infants in the concern group using the TSFI and the TIA and concluded that all 14 suffered from regulatory disorders. (2) Predictive validity: In a separate publication, the authors reported that 78 percent of infants identified by the ITSC as having problems were diagnosed with developmental or

behavioral problems at 3 years of age using standardized measures such as the Child Behavior Checklist.

Method of Scoring: The item responses are scored in the following manner: 0 points for "never or sometimes," 1 point for "past," and 2 points for "most times" (a self-calming item is the only item that is scored differently due to different response categories). The points are then summed for the entire checklist. The total score is then compared to a normal score range for the appropriate age group. A protocol sheet is available to assist in the scoring of the instrument.

Interpretability: The total checklist score is compared to the cutoff score for normal functioning infants and toddlers in the child's age group. Children whose scores fall at or above the cutoff score are considered to be at risk of having a regulatory disorder and further diagnosis is warranted.

Training Support: The manual includes case studies.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

DeGangi, Georgia A., Susan Poisson, Ruth Z. Sickel, and Andrea Santman Wiener. Infant/Toddler Symptom Checklist: A Screening Tool for Parents. San Antonio, TX: Therapy Skill Builders, Psychological Corporation, 1995.

LEITER INTERNATIONAL PERFORMANCE SCALE-REVISED (LEITER-R), 1997

Authors:

Gale H. Roid and Lucy J. Miller

Publisher:

Stoelting Co.

(630) 860-9700

www.stoeltingco.com

Initial Material Cost:

Complete Leiter-R Kit: \$850 (includes manual, 3 easel books, response cards, manipulatives, record forms, booklets, and carrying case)

Representativeness of Norming Sample:

The Leiter-R was standardized on 1,719 typical children and adolescents and 692 atypical children ages 2 years to 20 years, 11 months using a national stratification plan based on 1993 U.S. Census statistics for age, gender, and socioeconomic status. Nationally representative proportions of children who are Caucasian, Hispanic-American, African-American, Asian-American, and Native American were included.

Languages:

English

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

2 years to 20 years, 11 months

Personnel, Training, Administration, and Scoring Requirements:

The Leiter-R should be administered by a trained individual who has received supervised training and practice. It should be interpreted by someone with graduate training in psychological assessment. Administration times for the Leiter-R range from 25 to 40 minutes.

Summary:

Initial Material Cost: 3 (\$200 or higher)

Reliability: 3 (.65 or higher)

Validity: 3 (concurrent .5 or higher, no information or predictive)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a trained individual)

Description: The Leiter-R is an individually administered nonverbal test designed to assess cognitive functions in children and adolescents. It was developed to provide a reliable and valid nonverbal measure of intellectual ability, memory, and attention that could be used to assess children, adolescents, and young adults who could not be reliably and validly tested using traditional intelligence tests. The Leiter-R consists of two groupings of subtests: (1) the Visualization and Reasoning (VR) Battery (10 subtests), and (2) the Attention and Memory (AM) Battery (10 subtests). It also includes four social-emotional rating scales (Examiner, Parent, Self, and Teacher) that

provide behavioral observation information about the examinee. The majority of Leiter-R items require the child to move response cards into slots on the easel tray. Other items require arranging manipulatives (foam rubber shapes) and pointing to responses on the easel pictures. Starting points in the sub-tests are determined by the child's age (there are three age groups for administration of the Leiter-R: 2-5, 6-10, and 11-20).

Uses of Information: The Leiter-R can be used to help identify children with cognitive disabilities, to monitor small increments of improvements in cognitive abilities, and to

develop intervention strategies that address the identified disabilities. For initial screening purposes, four sub-tests in the VR Battery can be used to measure the child's global intellectual level as part of a battery of other tests and assessments. The full VR Battery (six subtests for children ages 2 to 5) can be used for identification, classification, and placement decisions. The AM Battery can be used for a comprehensive diagnostic assessment of attention and memory difficulties, neuropsychological evaluations, and evaluation of cognitive process deficits in learning disabilities or attention deficit disorders. Examiners have the option of using the VR and AM Batteries separately; however, the Batteries should be used together for a thorough cognitive assessment, particularly when it is expected that cognitive-process deficits in memory or attention are interfering with the accurate evaluation of global intellectual level.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): for children age 2, alphas ranged from .71 to .94 across the VR Battery subtests and, for children ages 2 to 3, the alphas ranged from .77 to .89 for the AM Battery subtests. For children ages 2 to 5, the alphas ranged from .71 to .90 for the AM Battery special diagnostic scale, and the reliability coefficients from .94 to .99 for the composite rating scale, and .87 to .93 for IQ and composite scores. (2) Test-retest reliability (with interval not reported): test-retest correlations ranged from .61 to .95 across the VR Battery subtest and composite scores and .86 to .94 across the examiner rating scales and

composite for children ages 2 to 5.¹ (3) Inter-rater reliability: no information reported in the manual.

Validity: (1) Concurrent validity: concurrent validity tests between the Leiter-R (Brief and Full Scale IQ) and the Wechsler Intelligence Scale for Children (WISC-III) (Performance and Full Scale IQ) on children ages 6 to 16 resulted in correlations of .85 and .86. The reported tests between Leiter-R Full Scale IQ scores with other cognitive tests showed correlations that ranged from .38 to .66. Tests for accuracy on children ages 2 to 20 showed that a cut-point of 70 on the Leiter-R Full Scale IQ score correctly classified more than 80 percent of children with cognitive delays. The classification accuracy of the Leiter-R for identifying giftedness was not as good, and the manual recommends that the Leiter-R never be used in isolation to identify giftedness. (2) Predictive validity: no information available.

Method of Scoring: The manual contains detailed scoring instructions. For most subtests, responses are scored as (0) Fail or incorrect, or (1) Pass or correct. Scoring criteria for each item are noted on the instruction page for each subtest. For some subtests, scoring requires counting the number of correct responses and the number of errors. Raw scores are typically obtained by summing correct responses. The raw scores on the subtests and rating scales are converted to scaled scores (with a mean of 10 and a standard deviation of 3) using a table provided in the manual. IQ scores are calculated from sums of

¹ No AM Battery retests were performed on children under the age of 6.

subtest scaled scores and converted to IQ standard scores (with a mean of 100 and standard deviation of 15) using a table in the manual. Composite scores can also be obtained for Fluid Reasoning, Fundamental Visualization, Spatial Visualization, Attention, and Memory. In addition, the raw scores for each subtest and IQ can be converted to growth-scale scores that define a child's domain of abilities in a metric that can reflect growth and be useful for treatment planning and measuring change over time.

Interpretability: Only persons with graduate training in psychological testing and statistics should interpret the results of the Leiter-R. The manual also cautions that IQ scores from the Leiter-R should never be used in isolation and should be evaluated in the context of a wide variety of information about the child. The manual includes an extensive discussion of the interpretation of Leiter-R results and provides case studies to demonstrate the interpretation of scores.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: The Leiter-R was specially developed to be used with children who could not be reliably and validly tested using traditional intelligence tests, including children with significant communication disorders, cognitive delays, English as a second language, hearing impairments, motor impairments, traumatic brain injury, attention-deficit disorder, and certain types of learning disabilities. The manual discusses adaptations to administration methods that may be needed for some children to establish that the child understands the nature of the task or to enable the child to communicate answers to test items within the capabilities they have. When such adaptations are made, growth-scale scores that do not rely on normative comparisons should be used.

Report Preparation Support: None described.

References:

Roid, Gale H., and Lucy J. Miller. Leiter International Performance Scale-Revised. Wood Dale, IL: Stoelting Co., 1995, 1997.

MACARTHUR-BATES COMMUNICATIVE DEVELOPMENT INVENTORIES (CDIs)—SECOND EDITION, 2007 (2011 Update)

Authors:

Larry Fenson, Virginia A. Marchman, Philip S. Dale, J. Steven Reznick, Donna Thal, and Elizabeth Bates

Publisher:

Paul H. Brookes Publishing Company
(800) 638-3775
<http://www.brookespublishing.com>

Instrument:

<http://www.brookespublishing.com/resource-center/screening-and-assessment/cdi/>

Initial Material Cost:

Complete Set of CDIs (Words and Gestures forms and Words and Sentences forms), User's Guide and Technical Manual, package of 20 of each form): \$99.95

Complete Set of CDIs and CDI III, User's Guide: \$121.95

CDI short forms available for purchase from authors: \$.35 each (http://www.sci.sdsu.edu/cdi/short_e.htm)

Representativeness of Norming Sample:¹

The updated 2007 norming sample for the CDIs (Words and Gestures, Words and Sentences) included 2,550 children (8 to 30 months of age) in New Haven, Connecticut; San Diego, California; Seattle, Washington; Dallas, Texas; Madison, Wisconsin; New Orleans, Louisiana; Providence, Rhode Island; and Storrs, Connecticut. The sample contained equivalent numbers of boys and girls; included children from black, Hispanic, and Asian racial/ethnic groups; and collected information about birth order, maternal

¹ The short forms were normed on a convenience sample of 1,379 primarily white infants and toddlers with parents who had high levels of education. Roughly 14 percent of infants and toddlers were exposed to a second language at home. Data were collected from New Haven, Connecticut; Seattle, Washington; and San Diego, California.

education, and exposure to languages other than English. Compared with 2000 U.S. Census Bureau information, the sample included a higher percentage of white children and mothers with college diplomas. The CDI-III was normed separately with 356 children age 30 to 37 months from a university subject pool. The sample included a similar number of boys and girls, and maternal education levels were higher than for U.S. Census population data. The Spanish speaking sample is described below (see Other Languages).

Languages:

English, Spanish

Type of Assessment:

Parent report

Age Range and Administration Interval:

8 to 37 months

Personnel, Training, Administration, and Scoring

Requirements: Administered by a clerical program staff member

Learning to administer CDIs requires less than one hour. They take 20 to 40 minutes for a parent to complete, and the short form takes 10 minutes. Hand scoring requires 20 to 30 minutes depending on age and language skill, although computer-based scoring programs are available.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (most .65 or higher)

Validity: 3 (most .5 or higher for concurrent; most .4 or higher for predictive)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (administered by someone with basic clerical skills and scored and interpreted by a highly trained individual)

Description: The MacArthur-Bates Communicative Development Inventories (CDIs), Second Edition is a parent report measure of early language skills of children between 8 and 37

months of age. The full tool consists of three inventories: (1) the CDI: Words and Gestures form (8 to 18 months); (2) the CDI: Words and Sentences form (16 to 30 months); and (3) the

CDI-III (30 to 37 months). Each form consists of several subtests. Short forms are available for the CDI: Vocabulary Checklist Level I (infant form) and Level II (toddler form).

The original edition of the MacArthur Communicative Development Inventories was published in 1992 and consisted of two inventories: Words and Gestures and Words and Sentences. The updated edition of the MacArthur-Bates CDIs adds the CDI-III, an extension to capture information on children age 30 to 37 months. The norming data for the CDI: Words and Gestures were expanded to include 17- and 18-month-olds. The updated addition expands the information on administration, interpretation, and scoring procedure options.

The CDI: Words and Gestures form assesses vocabulary production and comprehension and consists of two parts, each with several subtests (see Table 1 for subtest names and descriptions). Part I (Early Words) contains 429 items within four subtests and includes a 396-item Vocabulary Checklist divided into 19 semantic categories. Part II (Actions and Gestures) contains 63 items within five subtests.

The CDI: Words and Sentences form assesses increased vocabulary production and grammar acquisition and consists of two parts (see Table 1 for subtest names and descriptions). Part I (Words Children Use) contains 685 items within two subtests and includes a 680-item Vocabulary Checklist divided into 22 semantic categories. Part II (Sentences and Grammar) contains 113 items within six subtests. Parents must respond to a question about whether their child is combining words into sentences. If the response is no, two subtests (Examples or Complexity) need not be completed.

CDI-III is an extension of the CDIs for children 30 through 37 months of age. It is a short, single-sheet tool that measures expressive vocabulary and grammar. The first component features a 100-item Vocabulary Checklist (including 45 words from the CDI: Words and Sentences form and 55 new words). The second component consists of 13 questions about the child's word combinations (including 12 sentence pairs, of which 5 are drawn from the CDI: Words and Sentences form and 7 are new items). The third component consists of 12 questions, to be answered yes or no, that ask about various aspects of comprehension, semantics, and syntax.

Table 1

CDI SUBTESTS AND SCORES

Subtest (number of items)	Description	Score
CDI: Words and Gestures (Part I—Early Words)		
First Signs of Understanding (3 items)	General questions about early comprehension of familiar words and phrases	Not described
Phrases (28 items)	Comprehension of everyday phrases and routines	Phrases Understood
Starting to Talk (2 items)	Imitation and labeling	Not described
Vocabulary Checklist (396 items)	Checklist organized into 19 semantic categories; response options are understands or understands and says	Words Understood; Words Produced
CDI: Words and Gestures (Part II—Actions and Gestures)		
First Communicative Gestures (12 items)	Checklist of intentional gestures	Early Gestures; Total Gestures
Games and Routines (6 items)	Checklist of games the child plays, such as patty cake or peekaboo	Early Gestures; Total Gestures
Actions with Objects (17 items)	Checklist of actions the child is able to perform, such as brushing teeth, combing hair, or eating with spoon or fork	Later Gestures; Total Gestures
Pretending to Be a Parent (13 items)	Checklist of actions the child sometimes performs with stuffed animals or toys, such as putting them to bed or talking to them	Later Gestures; Total Gestures
Imitating Other Adult Actions (15 items)	Checklist of actions the child might try to imitate, such as cleaning with a broom, vacuuming, or washing dishes.	Later Gestures; Total Gestures
CDI: Words and Sentences (Part I—Words Children Use)		
Vocabulary Checklist (680 items)	Checklist of words the child can say, organized into 22 semantic categories	Words Produced
How Children Use Words (5 items)	Questions on the child's use of language for past, future, and absent objects and people	Not described
CDI: Words and Sentences (Part II—Sentences and Grammar)		
Word Endings/Part I (4 items)	Questions about the child's use of language to refer to past, future, and absent objects and people that differ from questions in the How Children Use Words subtest; for example, the subtest includes questions about how the child uses the possessive	Not described
Word Forms (25 items)	Checklist of irregular plural nouns and irregular past tense verbs	Word Forms
Word Endings/Part II (45 items)	Checklist of over-regularized nouns and verbs	Word Endings/Part II
Combining (1 item)	Question on whether the child can combine words into sentences	
Examples (1 item)	Request for parent to provide up to 3 of the longest sentences uttered by the child	Mean Length of the 3 Longest Sentences
Complexity (37 items)	Parents select one from a pair of sentences contrasting in complexity to indicate how their child currently speaks	Complexity

Source: Fenson et al. User's Guide and Technical Manual, 2007.

Other Languages: A Spanish version of the CDIs and short forms is called the Inventarios, published with the manual in 2003. The Inventarios were normed on more than 2,000 children in Mexico. The CDI-III is currently not available in Spanish. The CDI forms have also been adapted to 54 other country-specific languages such as Arabic, French, Finnish, and Mandarin (as spoken in Beijing, Singapore, and Taiwan).

Uses of Information: The CDIs may be used to screen for delays in language development and to identify problematic skills. The authors also note that the CDIs can help formulate intervention strategies and evaluate treatment outcomes.

Reliability:

1. Internal consistency reliability: Cronbach's alpha coefficients for scores from the CDI: Words and Gestures Total Gestures score had a reliability estimate of .88. The CDI: Words and Gestures Words Understood and Words Produced (on the Vocabulary Checklist) were .95 and .96, respectively. Within the Vocabulary Checklist, authors reported coefficients for semantic subcategories if they were below .70. The first semantic subcategory, Words about Time, had a coefficient of .65 for both Words Understood and Words Produced. The second semantic subcategory, Question Words, had coefficients of .68 and .56. In addition, scores for the Level I short form had a Cronbach's alpha of .97.

Cronbach's alpha coefficients for scores from the CDI: Words and Sentences form were .86 for the Words Produced scores and .95 for the Complexity scores (analyzed by using bound morphemes, functor words [i.e., words used for grammatical correctness such as "is," "the," or "a"], and complex sentences). Coefficients for semantic subcategories below .70 were Sound Effects and Animal Sounds (.65) and Connecting Words (.68). Scores for the Level II short forms A and B each demonstrated an internal consistency alpha of .99.

2. Test-retest reliability: for the CDI: Words and Gestures, authors analyzed 137 test pairs (number and age range of children not provided) with an average interval of 1.4 months between administrations through 12 months of age. Correlations were in the .80s for both Words Produced and Words Understood except at 12 months, when scores were correlated at .60. Authors noted that the decrease could be attributable to developmental transitions that occur at 12 to 13 months of age. Authors also correlated scores from the same forms by scale for longer administration intervals. On a sample of 62 children assessed at age 8 to 10 months and at age 14 to 17 months, the CDI: Words and Gestures correlation between scores by major inventory.

For CDI: Words and Sentences, vocabulary production correlated at .95 among 216 test pairs with an average interval of 1.4 months

between administrations. Authors also correlated scores from the same forms by scale for longer administration intervals. On a sample of 228 children age 16 to 24 months and age 22 to 30 months, the CDI: Words and Sentences correlated with itself at .71 for total vocabulary and at .62 for Complexity scores.

Authors also correlated scores between the CDI: Words and Gestures form (for use up to 18 months of age) and the CDI: Words and Sentences form (for use starting at 16 months of age) because some infants aged into the toddler group over the interval. The total vocabulary production correlation among 217 children between 10 and 16 months old and between 16 and 25 months old was .69.

The Level I short form yielded test-retest reliability estimates of .88 for Vocabulary Comprehension and .90 for Vocabulary Production based on a two-week interval and 50 infants (Hanson 1994). The test-retest reliability estimates for the Level II short forms A and B Vocabulary Production among 28 and 40 toddlers with a two-week interval were, respectively, .74 and .93.

3. Inter-rater reliability: not applicable.

Validity:

1. Content Validity: The authors drew items for each subtest from the developmental literature and used parent suggestions in response to earlier versions of the assessment. Major domains for infants include Words Understood, Words Produced, and Actions.

2. Concurrent validity: the authors correlated scores from the CDIs with several language measures. For CDI: Words and Gestures, the authors correlated scores from Words Produced with scores from the Language Sample NDW (Number of Different Words), the Preschool Language Scale (PLS-Revised), the Peabody Picture Vocabulary Test-III (PPVT-III), and the Reynell Developmental Language Scales (RDLS) Expressive subtest, with correlations ranging from .52 to .82 between scores. For the CDI: Words and Sentences, the authors used the same measures as above as well as the Bayley Scales of Infant Development (Second Edition) Expressive Language subtest, the Expressive One-Word Picture Vocabulary Test (EOWPVT), and the Sequenced Inventory of Communication Development-Revised (SICD-R), with correlations ranging from .40 to .88 between scores. Finally, for the CDI: Words and Gestures, the authors correlated scores from Words Understood to the Index of Productive Syntax, the PPVT-III, the RDLS Receptive, and the Language Sample NDW, with correlations ranging from .51 to .87.

Scores from the CDI-III Vocabulary Checklist correlated at .63 with the Preschool Language Scale-3 (PLS-3) total score, .58 with the PLS-3 Auditory Comprehension Score, and .47 with the PLS-3 Expressive Communication Score for a sample of 19 children (36 and 37 months of age). Correlations for scores of the Peabody Picture Vocabulary Test-Revised with the CDI-III Vocabulary Checklist, Sentences, and

Using Language subtests were .50, .45, and .63, respectively, for a sample of 22 children (36 to 39 months of age). Two separate studies correlated scores from the CDI-III with scores from the McCarthy Scales, with correlations ranging from .44 to .62 in the first study of 85 32- to 40-month-olds and from .52 to .56 in the second study of 113 3-year-olds.

The Early Head Start Family and Child Experiences Survey, or Baby FACES study, is a nationally representative, longitudinal study of Early Head Start programs. Unpublished findings show correlations between the short form Vocabulary Production and Vocabulary Comprehension (toddler form) scores as reported by parents and program staff with the Ages & Stages Questionnaires, 3rd Edition (ASQ-3) Communication subscale and Preschool Language Scale, Fourth Edition (PLS-4) (C. Vogel, Personal Communication, June 2011). Short form Vocabulary Production scores and the ASQ-3 Communication were correlated at .65 and .44 for parents' and staff's short form reports, respectively. Short form Vocabulary Comprehension scores as reported by staff and the ASQ-3 Communication score were correlated at .30. Short form Vocabulary Production scores reported by parents and staff correlated with the PLS-4 scores from .23 to .36.

3. Construct validity: the authors provided intercorrelations for the CDI: Words and Gestures and CDI: Words and Sentences

subtests. For the CDI: Words and Gestures, gestures were closely associated with vocabulary comprehension and moderately associated with vocabulary production. The Total Gestures scores correlated with Words Understood and Phrases Understood at .79 and .77, respectively, whereas Total Gestures scores correlated with Words Produced at .59. Words Understood and Phrases Understood were correlated at .80. Within the CDI: Words and Sentences, vocabulary production was highly correlated with grammar. Words Produced was correlated with Word Forms, Complexity, and Mean Length of the Three Longest Sentences at .83, .82, and .78, respectively.

Bias Analysis: No information available.

Methods of Scoring: Parents' response options vary by subtest. For example, for the Vocabulary Checklist, parents choose from response options of "understands" or "understands and says" words. In other parts, parents circle actions or gestures exhibited by their child or provide open-ended responses to questions about their child's longest utterances. Scoring may be performed manually or by using an automated process. The User's Guide and Technical Manual provide instructions for manual scoring and provide tables with percentile scores relative to age and gender. Scoring the inventories involves counting the number of marked items or affirmative responses by subtest. Thus, within an inventory, several subtests are combined to create a variety of composite scores (Table 1).

For the CDI: Words and Gestures form, the assessor determines five potential raw scores. In the Vocabulary Checklist of Part I, for each of the 22 semantic subcategories, items marked “understands” yield the Words Understood score, and those marked “understands and says” yield the Words Produced score (each has a maximum score of 396). Items marked “yes” in Part II, First Communicative Gestures and Games and Routines, are summed to yield the Early Gestures score (maximum score of 18) while those marked “yes” in the other Part II subtests of Actions with Objects, Pretending to Be a Parent, and Imitating Other Adult Actions make up the Later Gestures score (maximum score of 45). The Early Gestures and Later Gestures scores are summed for a Total Gestures score.

For the CDI: Words and Sentences form, the items within each subtest are also summed to provide five potential raw scores (Table 1). In Part I, the assessor calculates the Words Produced raw score by counting items marked as “says” in each of the 19 semantic subcategories of the Vocabulary Checklist (maximum score of 680). Additional subtests such as Word Endings/Part I, Word Forms, and Word Endings/Part II are individually scored by counting all items marked “sometimes” and “often” and computing total scores for each. The Examples subtest is scored by calculating the number of morphemes in each of the three example sentences and obtaining a mean length-of-three-longest-sentences score, instructions for which are provided in the User’s Guide and Technical Manual. The Complexity subtest is scored by counting the number of items

marked in the more complex of the two alternatives provided, yielding the Complexity score (maximum score of 37).

The CDI-III is scored by computing raw total scores for each of its three subtests: (1) Vocabulary Checklist (maximum score 100), (2) Sentences (maximum score 12), and (3) Using Language (maximum score 12). Assessors then convert the raw total scores into percentiles for comparison with the norming tables provided in the User’s Guide and Technical Manual. Tables in the User’s Guide permit the conversion of raw scores into gender- and age-specific percentile rankings.

The automated scoring process is completed by either machine scanning of original forms designed for use by image scanners (ranging in price from \$100 to \$3,000) or by the CDI Scoring Program’s entry of data at <http://www.sci.sdsu.edu/cdi/>. The program scores the English and Spanish forms as well as the short forms, calculates percentiles by using the raw scores, and generates child reports and parent letters.

Interpretability: The User’s Guide and Technical Manual provide detailed instructions for interpreting the results of the CDI: Words and Gestures and CDI: Words and Sentences. For these forms as well as for the CDI-III, the normed percentile ranking allows the child’s performance to be compared to that of other children in the child’s age group. For the CDI: Words and Sentences form, the authors noted a ceiling effect for the Animal Sounds category within the

Vocabulary Checklist but did not elaborate. Even though the measure may be self-administered and scored with use of the free downloadable automated program, the manual recommends that a clinician or a researcher interpret the results.

Training Support: No information available.

Adaptations/Special Instructions for Individuals with Disabilities: The manual cautions against using the CDIs with developmentally delayed children whose chronological age exceeds the upper limits of the inventory.

Report Preparation Support: An automated, free CDI scoring program is available at <http://www.sci.sdsu.edu/cdi/>. It generates child reports and parent letters for all forms and short forms of the inventory.

References:

Fenson, Larry, Virginia A. Marchman, Donna J. Thal, Philip S. Dale, J.S. Reznick, and Elizabeth Bates. *MacArthur-Bates Communicative Development Inventories (CDIs), Second Edition*. Baltimore: Paul H. Brookes Publishing Co., Inc., 2007.

Fenson, Larry, Virginia A. Marchman, Donna J. Thal, Philip S. Dale, J.S. Reznick, and Elizabeth Bates. *MacArthur-Bates Communicative Development Inventories, User's Guide and Technical Manual, Second Edition*. Baltimore: Paul H. Brookes Publishing Co., 2007.

Fenson, Larry, Steve Pethick, Connie Renda, and Jeffrey L. Cox. "Short-Form Versions of the MacArthur Communicative Development Inventories." *Applied Psycholinguistics*, vol. 21, no. 1, 2000, pp. 95-115.

Hanson, J.S. "A Validation of the Short-Form Versions of the MacArthur Communicative Development Inventories." Unpublished master's thesis. San Diego, CA: San Diego State University, 1994.

Jackson-Maldonado, Donna, Elizabeth Bates, and Donna J. Thal. *MacArthur Inventarios Del Desarrollo De Habilidades Comunicativas (Inventarios) (CDI-Spanish Edition)*. Baltimore: Paul H. Brookes Publishing Co., 2003.

MULLEN SCALES OF EARLY LEARNING, AGS EDITION, 1995

Authors:

Ellen M. Mullen

Publisher:

American Guidance Service, Inc.

(800) 328-2560

(800) 471-7220 (fax)

Customerservice@agsnet.com

Initial Material Cost:

Mullen Kit (birth to 68 months): \$636

Mullen Scales, Upgrade for Infant Mullen (birth to 39 months): \$282

Representativeness of Norming Sample:

Nationally representative sample of 1,849 children ages 2 days to 69 months with no known physical and mental disabilities and parents who spoke primarily English. Data on children in the northeast region was collected from 1981 to 1986 and from 1987 to 1989 for the south, west, north, and north central regions.

Languages:

English

Type of Assessment:

Direct child assessment (with some direct parent assessment)

Age Range and Administration Interval:

0 to 68 months

Personnel, Training, Administration, and Scoring Requirements:

For professionals with training or practical experience in the clinical assessment of infants and young children. The scales can be administered in approximately 15 minutes for 1-year-olds and 30 minutes for 3-year-olds.

Summary

Initial Material Cost: 3 (> \$200)

Reliability: 3 (.65 or higher)

Validity: 3 (.5 or higher concurrent validity)

Norming Sample Characteristics: 3 (nationally representative and mostly within the past 15 years)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Mullen Scales assess the cognitive functioning of young children from birth to 68 months. The assessment is based on the child's responses to activities prepared by the examiner. Believing that a global intellectual performance measure may mask uneven cognitive development, the Mullen Scales measure five distinct skills, Gross Motor and four "cognitive" skills—Fine Motor, Visual Reception, Receptive Language, and Expressive Language. The gross motor scale is administered to children from birth to 33 months and the four "cognitive" scales are administered to children from birth to 68 months. The "cognitive" scores can be summarized into an Early Learning Composite (ELC) score.

Uses of Information: Mullen scale scores can be used to identify children with special education needs who are eligible for further evaluation. The normative scores can also provide an objective means to identify weaknesses and strengths that underlie a child's learning style for the purpose of designing individualized instructional plans that capitalize on the child's strengths.

Reliability: (1) Internal consistency reliability: the median internal consistency split-half coefficients (Guilford's formula) for the five Mullen scales range from .75 to .83 and for the composite, .91. (2) Test-retest reliability (with a 1- to 2-week interval between tests): for the Gross

Motor Scale of the original Mullen scales, the correlation between tests was .96, and the median correlations on the “cognitive” scale were .84 (with a range of .82 to .85) for children ages 1 to 24 months and .76 (with a range of .71 to .79) for children ages 25 to 56 months. (3) Inter-rater reliability: correlations among raters ranged from .91 to .99 for age groups between 1 and 44 months.

Validity: (1) Concurrent validity: tests showed the Mullen scales to have stronger correlations with instruments that measured similar skills than those measuring different skills. The correlations of the Mullen “cognitive” scales with the Bayley Mental Development Index (MDI) were higher (.53 to .59) than their correlations with the Bayley Psychomotor Development Index (PDI; .21 to .52). The ELC also was more strongly correlated with the MDI (.70) than with the PDI (.43). Conversely, the Mullen Gross Motor scale was more strongly correlated with the Bayley PDI (.76) than with the MDI (.30). Similarly, the Mullen Receptive Language scale had a higher correlation with the Preschool Language Assessment Auditory Comprehension (.85) than with Verbal Ability (.72), while the converse was true with the Mullen Expressive Language Scale (.72 for auditory and .80 for verbal). Finally, the Mullen Fine Motor scale was strongly correlated with the Peabody Fine Motor Scale, across four age groups of children between the ages of 6 and 36 months (correlations ranged from .65 to .82). (2) Predictive validity: no information available.

Method of Scoring: The Item Administration Book provides instructions for scoring the items on the scales. Scoring is done on a record form containing a list of tasks or stimuli of possible responses for each assessment item. In most cases, the child receives a “1” for correct responses and “0” for incorrect responses. In some cases, the tester must sum the task scores to obtain the item scores. There are also cases where the item score can range from anywhere between 0 and 5. Scoring software (ASSIST) is available for purchase.

Interpretability: The raw scores for each scale can be converted into age-adjusted normalized scores. The four “cognitive” skills T score can be further converted into a normalized ELC score, which has a mean of 100 and a standard deviation of 15. In addition, the scores can be used obtain the child’s percentile rank and age equivalent score, the age at which the child’s raw score is the median score. The manual provides instructions for interpreting these scores, taking into account variables that may influence them. The ASSIST software program converts raw scores into the normalized scores and provides interpretative information.

Training Support: A training videotape can be purchased for \$104.95.

Adaptations/Special Instructions for Individuals with Disabilities: None.

Report Preparation Support: The manual provides three case studies as examples of how the Mullen Scales can be used and reported. The

Mullen ASSIST computer software program provides an optional narrative report.

References:

Mullen, Eileen M. Mullen Scales of Early Learning. Circle Pines, MN: American Guidance Services, Inc., 1995.

THE OUNCE SCALE, 2003 (2011 Update)

Authors:

Samuel J. Meisels, Dorothea B. Marsden, Amy Laura Dombro, Donna R. Weston, and Abigail M. Jewkes

Publisher:

Pearson Assessments
(800) 627-7271
<http://www.pearsonassessments.com>

Instrument:

<http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAaOunce&Mode=summary>

Initial Material Cost:

Ounce English Administrator Kit (includes the Standards, User's Guide, Reproducible Masters, and one of each Observation Record and Family Album): \$131.75

Ounce Online (includes online access to all Ounce products and introductory telephone training): \$10.95 per child for the first year (minimum of 10 children); in subsequent years, price depends on the number of participating children

Representativeness of Norming Sample:

No norming sample

Languages:

English, Spanish

Type of Assessment:

Observation and parent report

Age Range and Administration Interval:

Birth to 42 months; the Observation Record may be completed for infants and toddlers over the course of eight age levels ranging from 4 to 6 months in length, with the goal of completing several observations per question by the end of the age level.

Personnel, Training, Administration, and Scoring Requirements:

Parents, teachers, program staff, clinicians, and specialists may complete the Observation Record. Parents complete the Family Album, and program staff complete the Developmental Profiles.

Training for the Ounce Online is available by conference call or on site in a computer laboratory.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 2 (half at or under .65)

Validity: 2 (mostly under .5 for concurrent)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 2 (administered by someone with basic clerical skills)

Description: The Ounce Scale is an observation used to evaluate infant and toddler development from birth to 3.5 years. It is made up of three elements: (1) the Observation Record, a tool for tracking growth and development that provides caregivers with (a) a focus for observing and documenting children's daily behaviors and (b) data for making evaluations about development; (2) the Family Album, a structure that permits parents to learn about and record their child's development by writing down what

they see, using pictures and photographs, telling stories, and responding to observation questions similar to those in the caregiver's Observation Record; and (3) the Developmental Profiles, rating scales for caregivers and other staff to evaluate each child's observed development and progress as compared to developmental standards. The Ounce Scale is divided into eight age levels, each of which has its own Observation Record, Family Album, and Developmental Profile. From birth to 12 months, age levels span 4-month periods. From

12 to 42 months, age levels span 6-month periods. If an infant or toddler is within four weeks of moving into the next age level, the caregiver should use the Observation Record for the subsequent age group. For infants born prematurely, the caregiver should use the child's adjusted age until 2 years of age.

The Ounce Scale is organized around six developmental areas: Personal Connections, Feelings about Self, Relationships with Other Children, Understanding and Communicating, Exploration and Problem Solving, and Movement and Coordination. For each area, the Observation Record includes two to three questions about the child's behavior. The Developmental Profiles group the six developmental areas into four domains: social and emotional, communication and language, cognitive development, and physical development.

Other Languages: The print versions of the Family Albums and the reproducible masters are available in Spanish. Information on the norming sample, reliability, validity, and English language equivalence is unavailable.

Uses of Information: The Ounce Scale provides guidelines and standards for observing and interpreting young children's growth and behavior. Programs may use outcomes for accountability reporting, and parents and caregivers may use information to plan curricula and engage in enriching activities and experiences with infants and toddlers. The scale may not be used as a screening tool.

Reliability:

1. Internal consistency reliability: Meisels and colleagues conducted a validation and reliability study of the Ounce Scale (2010) with a sample of 287 Early Head Start (EHS) children evenly distributed across eight age groups four to six months apart. Slightly more than half of the children were male, and most were black or Hispanic. Eleven percent of the group was identified as having special needs. The study authors calculated a general scale of internal reliability by using 11 items shared across the eight age groups (with a different number of items for each age group), resulting in a Cronbach's alpha of .65. Reliability ranged from .19 to .89 when stratified by age group, with most groups showing reliabilities greater than .62.
2. Test-retest reliability: no information available.
3. Inter-rater reliability: not applicable. Meisels et al. (2010) explained that the nature of the test does not allow for truly independent observations because the Ounce Scale requires a high level of familiarity with the assessed child, and any co-teachers in the study combined their observations to produce Ounce ratings.

Validity:

1. Content validity: Developers used several methods to validate the content of the Ounce Scale. They reviewed over 100 books, instruments, and research materials and

created prototypes of the Ounce Scale, which were revised after informal review by program directors, researchers, and other field experts. Developers also solicited input from two expert panel meetings and revised the instrument based on the panels' feedback. Finally, they conducted two pilots and a year-long field test in more than a dozen early child care sites in six states and, to test and revise the scale, obtained feedback from program staff and families before finalizing the Ounce Scale.

2. **Concurrent validity:** the needs development ratings for the Ounce Developmental Profiles correlated with age-standardized criterion measures from the Preschool Language Scale-4 (PLS-4) and Bayley Scales of Infant Development-II (Bayley-II), with coefficients ranging from $-.28$ to $-.32$ (coefficients were negative because a higher score on the needs development portion of the Ounce indicates higher risk). The Ounce Developmental Profiles also correlated with the Ages and Stages Questionnaire: Social-Emotional (ASQ:SE) scale scores at $.47$ for all age groups combined (higher scores on the ASQ:SE indicate higher risk). Among subsamples based on children's ages and caregiver demographics, correlations were stronger for older children. The Developmental Profiles scores correlated well with the ASQ:SE for 4- and 24-month-olds and with the Bayley-II for 8-month-olds. Correlations were otherwise weak for younger groups.

Developers used Receiver Operating Characteristic Curve analysis to examine the sensitivity and specificity of the Ounce and the Bayley-II, PLS-4, and ASQ:SE. The Ounce and the other measures consistently identified children developing as expected based on specificity coefficients ($.75$ to $.77$ across measures) but less consistently identified children who need development based on sensitivity coefficients ($.63$ to $.77$ across measures).

Bias Analysis: No information available.

Methods of Scoring: Caregivers review the Observational Record and the Family Album and compare observations for the child to developmental standards, which provide examples of the behaviors that indicate each level of development. In the Developmental Profiles, they assess whether the child is developing typically by marking behaviors as "Developing as Expected" or "Needs Development" and recording comments in the available space.

Interpretability: If all the behaviors within an area of development are marked "Needs Development," the caregiver should consider providing additional support for the child in that area or pursuing additional assessment.

Training Support: The publisher offers training for the Ounce Online by conference call or on site in a computer laboratory. The training familiarizes users with the technology and is available at either a beginner or advanced level.

Adaptations/Special Instructions for Individuals with Disabilities: For children with developmental delays or disabilities, the caregiver may choose to use an Observation Record for a younger age in order to capture what the child is able to do rather than merely identifying areas that need development. Developers note that parents and professionals should be advised of the decision to use an earlier record.

Report Preparation Support: The caregiver uses the Observation Record, Family Album, and Developmental Profile ratings to complete a Summary Report of the child's development and to set goals. Ounce Online is an electronic record-keeping system for the Ounce Scale. It offers a Growth Report that allows teachers to track a child's progress, generates Summary Reports to share with families, and includes a Group Planning section that helps teachers identify and group children based on their needs (i.e., whether child behavior was rated as "developing as expected" or "needs development"). Ounce Online offers Outcomes Reports that may be used to fulfill EHS reporting requirements for demographic groups and/or progress within specific domains.

References:

Dichtelmiller, Margo L. *The Ounce Scale: Staff Development Manual*. New York: Pearson Early Learning, 2003.

Marsden, Dorothea B., Amy L. Dombro, and Margo L. Dichtelmiller. *The Ounce Scale User's Guide*. New York: Pearson Early Learning, 2003.

Meisels, Samuel J. "Development of the Ounce Scale." Available at [http://www.pearsoned.com/RESRPTS_FOR_POSTING/ASSESSMENT_RESEARCH/AR9.%20ounce_development.pdf].

Meisels, Samuel J., Amy L. Dombro, Dorothea B. Marsden, Donna R. Weston, and Abigail M. Jewkes. *The Ounce Scale Standards for the Developmental Profiles*. New York: Pearson Early Learning, 2003.

Meisels, Samuel J., Dorothea B. Marsden, Amy L. Dombro, Donna R. Weston, and Abigail M. Jewkes. "Ounce Online." Available at [<http://www.ounceonline.com/>]. 2003.

Meisels, Samuel J., Dorothea B. Marsden, Amy L. Dombro, Donna R. Weston, and Abigail M. Jewkes. *The Ounce Scale: An Infant/Toddler Observational Assessment*. New York: Pearson Early Learning, 2003.

Meisels, Samuel J., Xiaoli Wen, and Kristy Beachy-Quick. "Authentic Assessment for Infants and Toddlers: Exploring the Reliability and Validity of the Ounce Scale." *Applied Developmental Science*, vol. 14, no. 2, 2010, pp. 55-71.

PEABODY PICTURE VOCABULARY TEST, FOURTH EDITION (PPVT-4), 2007 (2011 Update)

Authors:

Lloyd M. Dunn and Douglas M. Dunn

Publisher:

Pearson Assessments

(800) 627-7271

<http://ags.pearsonassessments.com>

Instrument:

<http://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAa30700>

Initial Material Cost:

PPVT-4 Test Kit (includes Picture Easel, Manual, 25 Record Forms, and Carrying Case): \$220 for each form or \$407 for both forms

PPVT-4 ASSIST scoring CD-ROM Kit: \$150

Representativeness of Norming Sample:

The norming sample consisted of a stratified random sample of 3,540 individuals age 2 years, 6 months to over 90 years (between 100 and 200 each at one-year interval for ages 2 to 22 years) selected to match the U.S. population proportionately on gender, race/ethnicity, socioeconomic status, geographic region, and special education status. The sample was restricted to individuals proficient in English. A subsample of 2,003 students in kindergarten through grade 12, based on Census Bureau data on grade distribution, was used to establish grade norms. The

assessments were conducted from fall 2005 to spring 2006 at 320 sites nationwide.

Languages:

English

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

2.5 to 90 years

Personnel, Training, Administration, and Scoring Requirements:

Administered by a highly trained program staff member who should have a bachelor's degree with coursework in measurement, be thoroughly familiar with the test materials, and be properly trained in administering and scoring the test.

It takes 10 to 15 minutes to administer the 60 items to which average individuals respond.

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 3 (all .65 or higher)

Validity: 1 (validity information not available on children 0 to 3 years)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The PPVT-4 is an individually administered adaptive assessment designed to measure the receptive (auditory) vocabulary level for standard English. It is appropriate for people between the ages of 2 years, 6 months and 90 years and above. It has two parallel forms, Forms A and B, each with age-based training practice items and 228 test items grouped into 19 sets of 12 items,

with the sets arranged in order of increasing difficulty. During the assessment, the assessor presents a stimulus word orally with a set of four color pictures on an easel and asks the respondent to identify the picture that best represents the word's meaning. The assessor administers the item sets beginning at a predetermined age-appropriate start item until the basal and ceiling

sets are found. On average, individuals respond to 5 item sets. The basal set is set 1 or the first item set in which the respondent makes one or no errors. The ceiling set is either the first item set in which the respondent makes eight or more errors or the end of the assessment. By not requiring reading, writing, or speaking on the part of the respondent, the PPVT-4 is useful in assessing young children and may be used successfully with individuals with disabilities. Average administration time is 10 to 15 minutes (for 5 sets of 12 items or 60 items).

The previous version of the assessment, the PPVT-III (Dunn and Dunn 1997), is still used for research purposes. The publisher notes that the main updates to the PPVT-4 include colorized pictures with an increased balance of gender and racial diversity; more stimulus words, particularly at the floor and ceiling of the measure (easiest or most difficult); and growth scale value scoring, which may be used to measure a student's progress over time.

Other Languages: None.

Uses of Information: The PPVT-4 measures receptive vocabulary in standard English. The publisher reports that the PPVT-4 may also be used to (1) measure an individual's vocabulary growth and/or response to instruction; (2) diagnose reading difficulties; (3) measure language potential, non-readers' development or impairments or written- or expressive-language difficulties or other impairments (e.g., aphasia); (4) screen for verbal development; (5) establish rapport with a respondent as an initial component

in a larger battery of assessments; and (6) evaluate an English language learner's extent of vocabulary (though it cannot provide a normative score for comparison for such individuals).

Reliability:

1. Internal consistency reliability: the Spearman-Brown split-half reliability (within forms) ranged from .89 to .97 for Form A scores and from .91 to .97 for Form B scores for those age 2 years, 6 months to 24 years. Cronbach's alpha for the same age groups ranged from .93 to .98 for Form A scores and from .94 to .97 for Form B scores. Calculations of split-half reliabilities were based on separate analysis of the odd and even items in a Rasch analysis.
2. Test-retest reliability: correlation coefficients ranged from .91 to .94 between scores from the two administrations (with about a four-week interval) for age 2 to 14 years (n = 340).
3. Inter-rater reliability: no information available.

Validity:

1. Content validity: the pool of stimulus words appropriate for color picture illustration was culled mainly from Merriam-Webster's Collegiate Dictionary (2003) and several editions of Webster's New Collegiate Dictionary (1953, 1967, 1981) as well as from several other vocabulary or lexicographic resources. Stimulus words were grouped into 20 content categories. The manual details the decisions guiding word selection and picture development for stimulus words and

construction of the two parallel forms. The developers paid attention to colorization of the pictures for the PPVT-4 update, demonstrating sensitivity to demographic and disability issues. They gauged item difficulty by using classical and Rasch methods. The authors state that the word stimulus selection process provides qualitative evidence of the content validity of the PPVT-4 as a measure of standard American English receptive vocabulary.

2. Concurrent validity: studies correlated the PPVT-4 with four instruments that measure expressive vocabulary, language ability, and/or reading achievement: the Expressive Vocabulary Test, Second Edition (EVT-2); the Comprehensive Assessment of Spoken Language (CASL); the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4); and the Group Reading Assessment and Diagnostic Evaluation (GRADE). In addition, the PPVT-4 was correlated with the PPVT-III. Assessments were administered on the same day, except for the PPVT-III, which was given up to 11 days later. Sample sizes ranged from 110 to 425 participants, except for the EVT-2, which used the same norming sample of 3,540. Sample participants ranged in age from 2 to 24 years but typically were elementary or middle school students. Correlations between the PPVT-4 and the EVT-2 ranged across grades from .80 to .84. Correlation coefficients with CASL subtest scores ranged from .37 to .77. Correlations with the CELF-4 language subtest scores

ranged from .67 to .79. Correlations with the GRADE ranged across grades from .35 to .79 on the total test scores and from .27 to .79 on vocabulary and comprehension composite scores. Correlations with the PPVT-III scores ranged across grades from .79 to .83.

Developers examined the difference of PPVT-4 means among nine groups, including a giftedness group, a language delay and relevant disabilities group, and a non-clinical reference group from the norming sample (controlling for gender, race/ethnicity, and socioeconomic status). Results showed that all tests were statistically significant.

3. Predictive validity: six studies have been conducted with the PPVT-R (the second version of the PPVT) and later achievement, language, and other assessment results. For respondents in preschool through grade 5, correlations ranged from .14 to .66.

Bias Analysis: In pre-release trials, the developers conducted item bias analysis by using a Rasch-based method. They eliminated or revised and retested items that, during the first national tryouts, were determined to be biased with regard to gender, race/ethnicity, socioeconomic status, and region of the country; the publisher reports that it typically dropped from the assessment items determined to be biased during the second national trial.

Methods of Scoring: The raw score is obtained by subtracting the total number of errors in all sets from the number of the last item in the

individual's ceiling set. Raw scores may be converted into age- or grade-normative or developmental scores as well as into a non-normative growth scale value (GSV) score used to measure an individual's improvement over time. Normative scores include standard scores (mean = 100, standard deviation = 15), percentiles, normal curve equivalents (NCE), and stanines. Developmental scores include age and grade equivalent scores. Grade norms are available for kindergarten through grade 12. A series of tables permits the conversion of raw scores into GSV, normative, or developmental scores and corresponding confidence intervals. PPVT-4 ASSIST is a scoring software program available for purchase that scores, converts scores, interprets results, and provides progress and group score reports.

Interpretability: Only persons with formal training in psychological testing and statistics should interpret the results of the PPVT-4. The manual provides a brief description of each score and of its uses and limitations. Individuals may use GSV scores to compare PPVT-4 scores to the scores of previous PPVT administrations. Qualitative interpretations of incorrect answers may be conducted by using the classification of items by part of speech.

Training Support: Pearson Assessments offers in-service training and content presentations, some in person and some online.

Adaptations/Special Instructions for Individuals with Disabilities: Given that it requires no reading or writing, the PPVT-4 may

be administered to many groups with disabilities without any significant changes. The assessor's manual describes various modifications that may be made when administering the assessment to groups with various disabilities, specifically deaf or hard-of-hearing respondents. Interpretation of results from the hearing-impaired population should be tentative; an expert on deafness notes that the norms and other standards have not been determined for the hearing-impaired.

Report Preparation Support: The PPVT-4 ASSIST scoring software program provides individual and group reports of score summaries, diagnostic analyses, and progress reports for users conducting repeated administrations.

References:

Dunn, Lloyd M., and Douglas M. Dunn. *Peabody Picture Vocabulary Test-Fourth Edition Manual*. Minneapolis, MN: Wascana Limited Partnership, 2007.

Dunn, Lloyd M., and Leota M. Dunn. *Examiner's Manual for the Peabody Picture Vocabulary Test-Third Edition*. Circle Pines, MN: American Guidance Service, Inc., 1997.

Montgomery, Judy K. *The Bridge of Vocabulary: Evidence-Based Activities for Academic Success*. Minneapolis, MN: Pearson Assessments, 2007.

Pearson Assessments. "ASSIST: Automated System for Scoring and Interpreting Standardized Tests." Available at [<http://psychcorp.pearsonassessments.com/pai/ca/support/tchsupport/ppvtassist.htm>]. 2006.

PRESCHOOL LANGUAGE SCALE, FOURTH EDITION (PLS-4), 2002¹ (2011 Update)

Authors:

Irla Lee Zimmerman, Violette G. Steiner, Roberta Evatt
Pond

Publisher:

Pearson Assessments
(800) 627-7271
www.pearsonassessments.com

Instrument:

<http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8659-406>

Initial Material Cost:

PLS-4 English Value Pack with Manipulatives (includes Examiner's Manual, Picture Manual, 15 Record Forms, and 23 Manipulatives): \$305

PLS-4 Screening Test Kit (includes Stimulus Book/Test Manual with stimulus pages, technical information, administration and scoring directions, and 25 Record Forms for each age): \$142

Representativeness of Norming Sample:

The PLS-4 was normed in 2002 by using 2000 Census Bureau figures for children age birth through 6 years. The norming sample included 1,534 children (75 to 110 children for each age group, broken down by three-month intervals in the first year and six-month intervals thereafter). The sample included equal numbers of males and females, various ethnic minorities (39.1 percent), children with disabilities

¹ The PLS-5 was made available in the summer of 2011.

(13.2 percent), and bilingual speakers (3.4 percent) from 357 sites in 48 states. The Spanish speaking sample is described below (see Other Languages).

Languages:

English, Spanish

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

Birth to 6 years, 11 months. Administered at three-month intervals in the first year and six-month intervals for older children.

Personnel, Training, Administration, and Scoring

Requirements: Administered by a consultant or expert with clinical training. Paraprofessional staff may be trained to administer the assessment.

Requires 20 to 40 minutes to administer for infants birth to 11 months; 30 to 40 minutes for 1 to 3 years, 11 months; and 25 to 45 minutes for 4 to 6 years, 11 months. Screening test requires 5 to 10 minutes to administer.

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 3 (all .65 or higher)

Validity: 3 (all .5 or higher for concurrent)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by highly trained individual)

Description: The Preschool Language Scale, Fourth Edition (PLS-4) provides clinicians with a diagnostic instrument to evaluate language development and identify language disorders or delays among children from birth through age 6 years, 11 months. The PLS-4 is an individually administered assessment that tests receptive and

expressive language skills considered to be language precursors. The PLS-4 contains two clusters-Auditory Comprehension and Expressive Communication-and three supplemental measures (these measures are not incorporated into the test scores but are optional and provide additional information). The Auditory

Comprehension cluster measures a child's ability to be attentive and respond to stimuli in the environment and to comprehend basic vocabulary or gestures. The Expressive Communication cluster focuses on social communication, expressive language skills, and vocal development. Both clusters include subtests (called tasks), with 4 subtests for each three-month interval for age birth through 11 months and 12 receptive/expressive tasks for each six-month interval for age 1 through 6 years. In total, the PLS-4 contains 68 items, with 2 to 8 items in each subtest. It includes the use of manipulatives (such as a ball, rattle, cups, and crackers) and easel administration (a Picture Manual); the assessor uses the objects or pictures as prescribed to observe the child's reaction or response. The supplemental measures are the Articulation Screener (AS), Language Sample Checklist (LSC), and Caregiver Questionnaire (CQ). If a child between the ages of 2 years, 6 months and 6 years makes speech errors during testing, assessors may choose to administer the AS to determine if further articulation testing should be undertaken. The LSC evaluates the child's language skills in conversational speech. The CQ is completed by the parents of infants and toddlers age birth to 3 years to gather information on the child's communication behavior at home, the needs of the family, and areas that parents would like to see addressed.

The publisher's web site also notes a PLS-4 Screening Test that can be used to screen for a broad spectrum of speech and language skills in

young children. Paraprofessionals or teachers' aides may administer the PLS-4 Screening Test.

The PLS-4 is similar to the PLS-3 in its design, subscales, and overall skills assessed. The PLS-4 updated the standardization sample from 1980 Census Bureau figures. Developers improved the scale's age-appropriateness as well as the assessment floor and ceiling for two age groups: birth to 11 months and 5 through 6 years, 11 months. The Auditory Comprehension and Expressive Communication clusters for these age groups are now grouped into four two-month subtests.

Other Languages: A Spanish edition of the PLS-4 (Zimmerman, Steiner, and Pond, 2002) was normed in 2001 by using a different standardization sample of 1,188 Spanish-speaking children, of whom 51 percent were infants and toddlers age 0 through 3 years, 11 months. Eighty-one percent came from homes where Mexican Spanish was spoken. The Spanish edition with its own manual is available separately on the publisher's web site. The PLS-4 English and Spanish editions are similar in terms of concepts measured but are not equivalent forms of the assessment, and the scores on the two tests are not comparable.

Reliability (Spanish version):

1. Internal consistency: among infants and toddlers, Cronbach's coefficient alphas ranged from .51 to .89 for Auditory Comprehension, .65 to .89 for Expressive Communication, and .72 to .94 for Total Language Scores. Among

older children, coefficients ranged from .74 to .84 for Auditory Comprehension, .72 to .90 for Expressive Communication, and .82 to .92 for Total Language Scores.

2. Test-retest reliability: the sample included 188 children from the standardization sample age 2 years to 5 years, 11 months. Administration intervals ranged from 2 to 14 days (mean = 7.7 days). For toddlers age 2 years to 3 years, 11 months, correlations for Auditory Comprehension and Expressive Communication ranged from .73 to .84 and from .84 to .87 for the Total Language Score. For older children, they ranged from .77 to .86 for the clusters and .80 to .89 for the Total Language Score.
3. Inter-rater reliability: certain Expressive Communication items require open-ended responses. To measure the reliability of ratings on these items, 10 elementary school teachers were trained to score an average of six protocols from each age group from birth through 6 years, 11 months and had three weeks of experience in using the PLS-4 scoring rules. Inter-rater reliability was calculated for the full Expressive Communication cluster based on 100 protocols each scored by 2 out of 10 teachers, with a resulting correlation of .99.

Validity (Spanish version):

1. Content validity: developers conducted a literature review and a content review and gathered feedback from a survey administered

to clinicians who would potentially use the measure. Experts reviewed subtests for bias to ensure appropriateness for children from Spanish-speaking backgrounds, of varying socioeconomic status, and from different country regions. Developers also piloted the Spanish PLS-4 during a “tryout” in 15 states in 2000 to test the PLS-4 on 218 children, of whom 54 percent were infants and toddlers. Nineteen additional children with language disorders were included. A second bias review conducted during the pilot ensured that changes made to the measure were appropriate.

2. Concurrent validity: correlations with the previous version (PLS-3; 2- to 14-day intervals between administrations) were .67 for Auditory Comprehension scores and .71 for Expressive Communication scores for 104 children age 2 through 6 years. Developers noted that the lower scores for the PLS-3 resulted from the PLS-3 standardization sample’s reference to English speakers because of a shortage of Spanish speakers. Seventy 3- to 5-year-olds diagnosed with a language disorder were compared to 70 of their typically developing counterparts. Sensitivity and specificity of Total Language Scores were .91 and .63, respectively. Developers attributed the low specificity value to lower scores among typically developing children in the Spanish language standardization sample.

3. Construct validity: correlation between Auditory Comprehension and Expressive Communication across ages was .66.

Uses of Information: The PLS-4 may be used to assess language development and determine whether a child has a language disorder and, if so, whether the source of the disorder is auditory, expressive, or an overall problem. The resulting standard score and percentile ranks can also help determine the severity of the disorder and identify areas for in-depth testing before defining therapy goals.

Reliability:

1. Internal consistency reliability: for children from birth through 3 years, 11 months, Cronbach's alphas ranged from .67 to .94 for Auditory Comprehension scores, from .73 to .95 for Expressive Communication scores, and from .81 to .97 for the Total Language Score across subtests and ages (age groups were split by three-month intervals under age 1 and by six-month intervals thereafter). Among older children, coefficients ranged from .66 to .92 for Auditory Comprehension, .84 to .94 for Expressive Communication, and .86 to .96 overall.
2. Test-retest reliability: correlations between scores from two administrations (intervals of 2 to 14 days) were based on 218 2- to 6-year-olds from the standardization sample. For 2-through 3-year-olds ($n = 102$), correlations for Auditory Comprehension ranged from .87 to .95, from .82 to .94 for Expressive

Communication, and from .90 to .97 for the Total Language Score. For older children, they ranged from .83 to .95 for subtests and from .80 to .89 for the Total Language Score.

Correlations between scores from two administrations (intervals of 2 to 14 days) ranged from 0.85 to 0.91 for Auditory Comprehension, 0.82 to 0.94 for Expressive Communication, and 0.90 to 0.97 for total scores across age groups (divided by five-month intervals) based on 218 2- to 6-year-olds from the standardization sample.

3. Inter-rater reliability: fifteen elementary school teachers were trained to score open-ended Expressive Communication items. They scored an average of six protocols from each age group from birth through 6 years, 11 months and had three weeks of experience in using the PLS-4 scoring rules. Inter-rater reliability was calculated for the full Expressive Communication cluster based on 100 protocols each scored by 2 out of 15 scorers, with a resulting correlation of .99.

Validity:

1. Content validity: developers measured relevance and coverage of the content based on a literature review, a user survey, and content reviews. The PLS-3 tasks were modified by using research data, and speech-language pathologists developed new tasks and items. Developers used a "tryout" or pilot to test the PLS-4 on a national sample of 661 children at 227 sites in 46 states. The tryout

phase consisted of 229 subtests (tasks) for each age group as well as tasks from the PLS-3. To determine any aspects of the subtests that may affect children inappropriately, the pilot tested an additional 53 children with language disorders. Developers then revised or deleted specific subtests and their subitems once the results were collected from the bias review, statistical analyses, and examiner feedback.

2. Concurrent validity: correlations with the previous version (PLS-3; 2- to 14-day intervals between administrations) were .65 for Auditory Comprehension scores and .79 for Expressive Communication scores for 104 2- through 6-year-olds. Thirty-seven infants birth to 11 months were administered the PLS-4 and the language strand of the Denver II in counterbalanced order on the same day. All the infants scored within the “normal” range on the Denver II and within one standard deviation of the mean on PLS-4.
3. Construct validity: correlation between Auditory Comprehension and Expressive Communication across ages was .74.

Bias Analysis: New or modified tasks were submitted to a panel of experts for two bias reviews to determine appropriateness for children from varied backgrounds (for example, socioeconomic status, ethnicity, and geographic region).

Methods of Scoring: The assessor records the source/type of response to each item and marks it

as correct or incorrect. To obtain a raw score, the assessor sums the items with a 1 (a correct response). A Total Language Score as well as Auditory Comprehension and Expressive Communication scores may be obtained. Raw scores may then be converted to standard scores, percentile ranks, and age equivalents.

Interpretability: The Examiner’s Manual provides detailed information on the interpretation of scores as related to determining the severity of the disorder and the need for intervention. Using the task analyses (PLS-4 Checklist and Profile), a clinician may evaluate the child’s strengths, emerging skills, and deficits. The Checklist groups PLS-4 subtests by age; the Profile groups subtests by the type of language skill tested.

Training Support: No information available.

Adaptations/Special Instructions for Individuals with Disabilities: The Examiner’s Manual provides special instructions on administration for children with autism and other severe developmental delays, children with physical impairments (such as hearing and vision), and children who use sign language. For example, the instructions call for different start points depending on level of impairment or disability, splitting up administration of the assessment into short sessions, removing distracting elements from the room, or using gestures/pointing. While the Examiner’s Manual provides no special instructions for children who are English Language Learners (ELL), it offers

some instructions for children from “non-mainstream” cultures.

Report Preparation Support: Assessors may use the score graph on the front page of the Record Form to provide caregivers with a visual representation of a child’s scores. The Clinician’s Worksheet may be used to help summarize information from the assessment. The worksheet has three sections: assessment results, followup for child and family, and outcomes of additional assessment and training. The worksheet is designed to include information relevant to the development of the Individualized Family Service Plan.

References:

Flowerday, Terri. “Review of Preschool Language Scale, Fourth Edition.” In *The Sixteenth Mental Measurements Yearbook*, edited by Robert A. Spies and Barbara S. Plake. Lincoln, NE: The Buros Institute of Mental Measurements, 2005.

Suen, Hoi K. “Review of Preschool Language Scale, Fourth Edition.” In *The Sixteenth Mental Measurements Yearbook*, edited by Robert A. Spies

and Barbara S. Plake. Lincoln, NE: The Buros Institute of Mental Measurements, 2005.

Zimmerman, Irla L., Violette G. Steiner, and Roberta E. Pond. *PLS-3: Preschool Language Scale-3*. San Antonio, TX: The Psychological Corporation, 1992.

Zimmerman, Irla L., Violette G. Steiner, and Roberta E. Pond. *Preschool Language Scale-4: Picture Manual*. San Antonio, TX: Harcourt Assessment, Inc., 2002.

Zimmerman, Irla L., Violette G. Steiner, and Roberta E. Pond. *Preschool Language Scale, Fourth Edition*. San Antonio, TX: Harcourt Assessment, 2002.

Zimmerman, Irla L., Violette G. Steiner, and Roberta E. Pond. *Preschool Language Scale, Fourth Edition, Examiner’s Manual*. San Antonio, TX: Harcourt Assessment, 2002.

Zimmerman, Irla L., Violette G. Steiner, and Roberta E. Pond. *Preschool Language Scale, Fourth Edition (PLS-4) Spanish Edition*. San Antonio, TX: Harcourt Assessment, Inc., 2002.

RECEPTIVE-EXPRESSIVE EMERGENT LANGUAGE TEST - SECOND EDITION (REEL-2), 1991

Authors:

Kenneth R. Bzoch and Richard League

Publisher:

PRO-ED

(800) 897-3202

www.proedinc.com

Initial Material Cost:

REEL-2 Complete Kit: \$86 (includes Examiner's Manual and 25 Profile/Test Forms)

Representativeness of Norming Sample:

Not representative, norming sample included only white infants.

Languages:

English

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

0 to 3 years

Personnel, Training, Administration, and Scoring

Requirements: It is possible for a consultant or expert with clinical training to learn to administer the REEL-2 in two hours. The manual suggests that a trainee be observed by an individual with REEL-2 experience. Administration time per child is 10 to 15 minutes; scoring time is 5 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (.65 or higher)

Validity: 1 (not available)

Norming Sample Characteristics: 2 (only white infants included)

Ease of Administration and Scoring: 3 (administered and scored by a clinician or expert with clinical training)

Description: The REEL-2 assesses communication behaviors (receptive and expressive) of infants and toddlers from birth to 3 years of age. The format is a 132-item checklist of language milestones, with 3 items contained in each of the 22 age intervals (age intervals vary from 1 to 3 months depending on the chronological age of the child). The examiner completes the checklist based on information provided by the child's caregiver. Typically, more than three items need to be administered to obtain the ceiling age interval for the child (further details are provided in the "Method of Scoring" section below). The REEL-2 uses caregiver report to identify any major language problems. Scores derived from the REEL-2 include an Expressive Language Quotient, a Receptive Language

Quotient and a Language Quotient. The REEL-2 is undergoing revision, and the publisher expects the third edition to be released in late 2003 or mid-2004.

Uses of Information: The primary uses of the REEL-2 are to (1) provide descriptions of the developmental status of young children in the language area, (2) assist with setting intervention goals, and (3) serve as a screening instrument for medically and environmentally at-risk populations. The REEL-2 is also used to determine the extent to which interventions have changed the language status of individual children.

Reliability: (1) Internal consistency reliability (Cronbach's alpha):.98 for 0 to 11 months, .97 for 12 to 23 months, and .95 for 24 to 36 months, with

an average of .97. (2) Test-retest reliability: .90 to 1.0.

Validity: (1) Concurrent validity: studies showed that the REEL-2 relates well to normal expected levels of functioning (Eich, 1971). However, studies were completed on small samples with similar characteristics. (2) Predictive validity: no information available.

Method of Scoring: The examiner uses the child's chronological age to determine the age interval from which the questions should first be asked. Behaviors observed for each item are scored as either typical of the child (+), emergent (+-), or never observed (-). Scores are computed by summing the (+) responses. The examiner then determines the ceiling interval, which is the highest age interval receiving at least two (+) item scores. The results of the Receptive Language Age and the Expressive Language Age are combined to form the Combined Language Age. Then, each of these scores (the Receptive Language Age, Expressive Language Age, and the Combined Language Age) are respectively divided by the child's chronological age and multiplied by 100 to obtain a Receptive Quotient, an Expressive Quotient, and a Language Quotient.

Interpretability: The Examiner's Manual includes average scores for the limited norming sample of children used to determine the validity of the REEL-2, but does not contain any overall score percentiles or cutoffs to guide interpretation of the REEL-2 scores. However, it does contain several interpretation guidelines, including the

relationship between Receptive Language skills and hearing and nervous system disorders, as well as the relationship between delays in Receptive Language versus Expressive Language. It also includes suggestions for intervention strategies related to language stimulation.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: The Examiner's Manual contains general suggestions on how to present reports/recommendations to parents: (1) the interviewer should provide parents/caregivers with the REEL-2 manual and other relevant booklets; (2) information about the score should be accompanied by the interviewer's explanation, which could include other interpretations and recommendations for intervention programs; and (3) the interviewer should avoid using technical language to explain the REEL-2 scores.

References:

Bzoch, Kenneth, and Richard League. Receptive-Expressive Emergent Language Test: A Method for Assessing the Language Skills of Infants, Second Edition. Austin: Pro-Ed, 1991.

Bzoch, Kenneth, and Richard League. Receptive-Expressive Emergent Language Test: A Method for Assessing the Language Skills of Infants, Second Edition: Examiner's Manual. Austin: Pro-Ed, 1991.

Eich, W.F. A Validation Study for the REEL. (Unpublished manuscript), 1971.

THE ROSSETTI INFANT-TODDLER LANGUAGE SCALE: A MEASURE OF COMMUNICATION AND INTERACTION, 1990

Authors:

Louis Rossetti

Publisher:

LinguiSystems

(800) 776-4332

www.linguisystems.com

Initial Material Cost:

Complete kit: \$70 (includes Examiner's Manual plus 1 package of forms, enough for 10 assessments)

Representativeness of Norming Sample:

Not representative. Tool is based on the research and experience of the author.

Languages:

English

Type of Assessment:

Because responses are elicited, observed, and reported, this measure is a combination of direct child assessment, observation, and parent report.

Age Range and Administration Interval:

Birth to 3 years; administered based on child's age in 3-month groupings.

Personnel, Training, Administration, and Scoring

Requirements: Designed for someone with a background in clinical assessment of child development and language. Training requires familiarizing oneself with the manual and questionnaires. Administration of the test will vary depending on whether or not the Parent Questionnaire was filled out and mailed prior to the test day as well as the age of the child (approximately 45 minutes for each age interval). Scoring is done concurrently with the test.

Summary:

Initial Material Cost: 1 (< 100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a clinician or expert with clinical training)

Description: The Rossetti Infant-Toddler Language Scale assesses the language skills of children from birth through 3 years of age. The scale assesses preverbal and verbal areas of communication and interaction through direct observation and caregiver report. Areas assessed include (1) Interaction-Attachment (relationship between the caregiver and the infant), (2) Pragmatics (the way language is used to communicate and affect others), (3) Gestures, (4) Play (both individual and interactive), (5) Language Comprehension, and (6) Language Expression. The test consists of a parent questionnaire and an examiner's evaluation form

that address each of the 6 areas. The parent questionnaire allows the examiner to familiarize himself with the developmental concerns expressed by the parent and helps to determine the age level at which testing should begin. The examiner's job is to establish both a baseline and ceiling developmental age by observing, eliciting, or using a caregiver's report of various behaviors listed in each of the six developmental areas.

Uses of Information: The primary use of the Rossetti Infant-Toddler Language Scale is the early detection of language delays in infants and toddlers. When assessment results indicate a child

needs early language intervention, a therapy program is developed with specific goals.

Reliability: This measure has not been standardized, and there is no statistical information on it.

Validity: This measure has not been standardized, and there is no statistical information on it.

Method of Scoring: If a parent questionnaire is completed, the description of current skills helps to determine the age level at which testing should begin. Otherwise, the examiner uses the child's chronological age to determine the age interval from which the questions should first be asked. A baseline and ceiling level of performance is established in each of the six developmental areas. To establish a baseline level (all items are mastered in the developmental area), the assessment is begun at six months below the child's chronological age or suspected developmental level. Once a baseline level is established, testing proceeds forward until the child fails all items for a developmental area at a particular age range (ceiling level). Items are considered "passed" if the behavior in question is noted in one of the 3 following ways: (1) Observe (O); (2) Elicit (E); (3) Report (R).

Interpretability: An individual baseline and ceiling age level for each of the six developmental areas may be reported in order to determine the child's relative areas of strengths or weaknesses. In addition, an examiner can compute a global baseline and ceiling age level, which is the oldest

age level at which the child mastered all items across all developmental areas. The global basal and ceiling can provide information about the child's performance ability to compare to his/her chronological or adjusted age.

Training Support: The examiner is free to call the publisher with questions related to the administering of the test. In addition, the manual has tips on how to elicit responses or when and where to look for them.

Adaptations/Special Instructions for Individuals with Disabilities: None

Report Preparation Support: The Examiner's Manual contains general suggestions on how to present reports and recommendations to parents. The examiner should (1) remain cautious of providing long-term predictions about the child's potential and needs; (2) remain sensitive to the amount of detail that is offered during the initial conference; (3) schedule a second conference to go over the results in greater detail; (4) actively involve the caregivers in the conference as soon as possible and ask for feedback from the caregiver; and (5) present programming recommendations as options rather than requirements.

References:

Rossetti, Louis. *The Rossetti Infant-Toddler Language Scale: A Measure of Communication and Interaction*. East Moline, IL: LinguiSystems, 1990.

TEMPERAMENT AND ATYPICAL BEHAVIOR SCALE (TABS) – EARLY CHILDHOOD INDICATORS OF DEVELOPMENTAL DYSFUNCTION, 1999

Authors:

John T. Neisworth, Stephen J. Bagnato, John Salvia, and Frances M. Hunt.

Publisher:

Paul H. Brookes Publishing Co.
(800) 638-3775
www.brookespublishing.com

Initial Material Cost:

Complete TABS System: \$85 (includes the manual, a pad of Screeners, and a packet of Assessment Tools)
Manual only: \$40

Representativeness of Norming Sample:

621 children of diverse racial and ethnic backgrounds between ages of 11 and 71 months; 52 percent were 2 years old or younger. Various geographic and socioeconomic conditions. Two separate samples were tested: children with disabilities and those without disabilities. Precisely representative normative samples were not necessary, because the occurrence of atypical behaviors in infants and young children is presumed to be unrelated to various demographic factors.

Languages:

English

Type of Assessment:

Parent report (a professional familiar with child's behavior may also report)

Age Range and Administration Interval:

11 to 71 months

Personnel, Training, Administration, and Scoring

Requirements: Child's parent(s) or a professional familiar with child's behavior can administer the TABS Screener and Assessment Tool. Administration time per child is 5 minutes for the Screener and 15 minutes for the Assessment Tool. Total time for assessment, scoring and interpretation is 30 minutes for each child.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (.65 or higher)

Validity: 1 (none described); content and construct validity are promising; concurrent and predictive validity not mentioned.

Norming Sample Characteristics: 3 (normed within past 15 years, representative of population program works with).

Ease of Administration and Scoring: 2 (administered by parent or professional who is familiar with child's daily behavior)

Description: The TABS assesses atypical temperament and self-regulatory behaviors among infants and young children ages 11 to 71 months that may put them at risk for developmental delay. There are two components to the assessment: (1) a Screener, which is a one-page, 15-item checklist with "Yes" or "No" responses that is used to determine whether further assessment is needed and (2) an Assessment Tool, which is a 55-item checklist with "Yes," "No" and "Need help" responses, used

to obtain a total raw score, or the Temperament and Regulatory Index (TRI). Raw scores are also provided for the four subtests on the following behaviors: detached, hyper-sensitive/active, underreactive, and dysregulated.

Uses of Information: The primary uses of the TABS are: (1) to serve as a screening tool and determine eligibility for particular services, (2) to identify serious developmentally dysfunctional behaviors early and intervene to reduce their occurrence, (3) to provide assistance to parents in

managing atypical behaviors, (4) to plan programs for education, treatment, and intervention, and (5) to evaluate program impacts and conduct research.

Reliability: (1) Internal consistency (Cronbach's alpha) for the TABS Screener is .83. (2) The corrected split-half reliability for the TABS is .95 for children with disabilities and for the pooled samples and .88 for children not at risk.

Validity: Content validity has been ensured in two ways: (1) TABS items have been developed based on an extensive literature review that identified behaviors related to a number of infant and child disorders, and (2) the four factors underlying the TABS have shown promising results when tested statistically. There was no discussion of concurrent or predictive validity.

Method of Scoring: In administering the Screener and the Assessment Tool, the examiner will have checked a "Yes" if a certain behavior is a current or recent problem and a "No" if the behavior is not a problem or does not apply because of the child's age. In scoring the Screener, the examiner adds up the items that are checked "Yes" and that sum becomes the raw score, which serves as an estimate for the Temperament and Regulatory Index (TRI). Children who have raw scores of 1 or higher on the Screener should be assessed using the Assessment Tool, which more accurately calculates the TRI. There are three types of scores that can be derived from the TABS Assessment Tool: percentiles, standard scores, and normalized standard scores. Typically, the

percentiles are the most practical tool for interpreting TABS results for parents, as well as for determining educational and treatment plans for children.

Interpretability: The authors suggest that the child's TRI score should be reported, interpreted and used for decision-making purposes, as the subtest scores tend to be less reliable than the complete TRI. A percentile means that a child has scored equal to or better than the respective percentage of children in the normative sample. For example, a percentile of 70 indicates that a child has a raw score equal to or better than 70 percent of the sample on the TRI. The cutoff point for the TRI is 10. This means that children who score 10 or higher can most likely be classified as having atypical development, meaning they have difficulty with their temperament and self-regulation. Children with disabilities are more likely to earn higher raw scores (6+) while those without disabilities are more likely to earn lower raw scores (0 to 4). A child is at risk for atypical temperament and/or self-regulation if s/he has a TRI score between 5 and 9, because children with disabilities earn those scores equal to or more often than their non-disabled peers.

Training Support: "Brookes on Location" professional development seminar, Using TABS to Identify Early Atypical Behavior is available through publisher.

Adaptations/Special Instructions for Individuals with Disabilities: None described, but the measure was normed with children with and without disabilities.

Report Preparation Support: Chapter 6 describes an early intervention program to address challenging behavior that was developed using research editions of TABS. A case study is also provided in this chapter to demonstrate how TABS might be used in an early intervention program of this nature.

References:

Neiswroth, John, Stephen Bagnato, John Salvia, and Frances Hunt. TABS Manual for the Temperament and Atypical Behavior Scale: Early Child Indicators of Developmental Dysfunction. Baltimore: Paul H. Brookes Publishing Co., 1999.

TEST DE VOCABULARIO EN IMAGENES PEABODY (TVIP), 1986

Authors:

Lloyd M. Dunn, Eligio R. Padilla, Delia E. Lugo, and Leota M. Dunn

Publisher:

American Guidance Services
(800) 328-2560
www.agsnet.com

Initial Material Cost:

TVIP Test Kit: \$119 (includes test easel, English or Spanish manual, and 25 record forms)
TVIP Manual: \$40
25 Record Forms: \$27

Representativeness of Norming Sample:

The norming sample consisted of monolingual, Spanish-speaking students in Latin America. Testing in Mexico took place between September 1981 and November 1982 and included 1,219 children from the public schools of Mexico with 20% coming from Mexico City. Testing in Puerto Rico took place between September 1982 and February 1983 and included 1,488 children from Puerto Rico with 62.2% from the San Juan metropolitan area. To correct for unevenness of socioeconomic status (SES) representation, a weighting system was used to increase or decrease the contributions of each individual's score at each age, so as to fit the SES ratios established by the U.S. census statistics.

Languages:

Spanish (the Peabody Picture Vocabulary Test is the English Version of this test)

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

2 1/2 to 18 years

Personnel, Training, Administration, and Scoring Requirements:

Although formal training in psychometrics is not required, the examiners should be thoroughly familiar with the test materials and well-trained in administering and scoring the test. It is extremely important that the examiner be proficient in correctly pronouncing each stimulus word. It takes 10 to 15 minutes to administer.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher for internal consistency, test-retest and inter-rater not described)

Validity: 1 (validity information not available on children from 0-3)

Norming Sample Characteristics: 2 (older than 15 years, not nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a trained individual)

Description: The TVIP is based on the Peabody Picture Vocabulary Test-Revised (PPVT-R) and includes 125 translated items to assess the vocabulary of Spanish-speaking and bilingual children. Items were selected through item analysis for their universality and appropriateness to Spanish-speaking communities. In the test, the examiner orally presents a stimulus word with a set of pictures and the test taker is asked to select the picture that best represents the word's meaning. The examiner administers the items until the child's "basal" and "ceiling" are found.

The basal is the highest set of eight consecutive correct responses and the ceiling is the lowest set of eight consecutive responses containing six errors. Because it requires no reading or writing and is easy to administer, the TVIP is useful in assessing older toddlers and preschool children and is fair to persons with written-language problems and disabilities such as autism, withdrawn personalities, psychotic symptoms, severe cerebral palsy, and moderate visual disabilities.

Uses of Information: The TVIP can be used to measure a child's receptive or hearing vocabulary of single Spanish words. It may be used as a screening test of verbal ability or verbal intelligence when Spanish is the language of the home and community into which the child was born and when Spanish is, and has been, the primary language of instruction in the child's program. It may also be used as an achievement test showing the extent of Spanish vocabulary acquisition.

Reliability: (1) Internal consistency reliability (split-half reliability): the median correlation coefficient, corrected using the Spearman-Brown formula, was .93. For age 2_ to 3, the coefficient was .80. (2) Test-retest reliability: no information available. (3) Inter-rater reliability: no information available.

Validity: (1) Concurrent validity: Correlations ranged from .25 to .59 between scores on the TVIP and the Kaufmann-ABC Global Scales and from .28 to .69 between the TVIP and the Kaufman-ABC Achievement Scale Subtests among children ages 3 to 6. The correlation between TVIP and the Habilidad General Ability test was .44 among children attending an urban private school in Puerto Rico. (2) Predictive validity: no information available.

Method of Scoring: The raw score is obtained by subtracting the total number of errors between the basal and ceiling sets from the number associated with the highest item in the ceiling set. Using several tables, raw scores can be converted

into different types of age-adjusted standardized scores using Mexican norms, Puerto Rican norms, or norms for a composite group. One table converts the raw score into a standard score. A second table converts the standard score into percentile rank, decile, and stanine. Finally, a third table converts the examinee's raw score into the age equivalent performance.

Interpretability: Only persons with graduate training in psychological testing and statistics who are familiar with the research literature on the language and cognitive development of Hispanic children should interpret the results of the TVIP.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: Because no reading or writing is required, the TVIP can be administered to many groups with disabilities without any changes.

Report Preparation Support: None described.

References:

Dunn, Lloyd M., Eligio R. Padilla, Delia E. Lugo, and Leota M. Dunn. Examiner's Manual for the Test De Vocabulario en Imagenes Peabody: Adaptacion Hispanoamericana. Circle Pines, MN: American Guidance Service, Inc., 1986.

Dunn, Lloyd M., and Leota M. Dunn. Peabody Picture Vocabulary Test-Revised. Circle Pines, MN: American Guidance Service, Inc., 1981.

TEST OF PRESCHOOL EARLY LITERACY (TOPEL), 2007 (2011 Update)

Authors:

Christopher J. Lonigan, Richard K. Wagner, Joseph K. Torgesen, and Carol A. Rashotte

Publisher:

PRO-ED, Inc.
(800) 897-3202
<http://www.proedinc.com>

Instrument:

<http://www.proedinc.com/customer/productView.aspx?ID=4020&SearchWord=topel>

Initial Material Cost:

Complete kit (Examiner's Manual, Picture Book, and 25 Record Booklets in a storage box): \$232

Representativeness of Norming Sample:

The norming sample consisted of 842 children, including 212 3-year-olds, 313 4-year-olds, and 317 5-year-olds from 12 states tested in 2004. This convenience sample was based on assessors in the PRO-ED customer files who tested 20 children each. The norming sample closely approximates the U.S. population, based on the 2001 Census Bureau, for region, gender, race, Hispanic ethnicity, family income, parent education attainment, and exceptionality status (such as a language disorder, attention-deficit/hyperactivity disorder, or a disability). Developers present age-stratified demographic variables that parallel national estimates.

Languages:

English

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

3 through 5 years

Personnel, Training, Administration, and Scoring Requirements:

Administered by a highly trained program staff member. The assessor should have knowledge of and experience in test administration, test scoring, and interpretation of norm-referenced results. Self-training involves reading the manual and practice by giving five trial administrations.

The TOPEL takes approximately 30 minutes to administer.

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 3 (all .65 or higher).

Validity: 3 (mostly .5 or higher for concurrent)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The TOPEL is an individually administered adaptive assessment of early literacy normed for 3- through 5-year-olds. The assessment consists of 98 items with three subtests: Print Knowledge (36 items), Definitional Vocabulary (35 items), and Phonological Awareness (27 items). The Print Knowledge subtest measures written language conventions

and alphabet knowledge. The child points to, identifies, or says the sounds associated with letters, words, and aspects of print. The Definitional Vocabulary subtest measures a child's single-word oral vocabulary and definitional vocabulary. The child identifies a picture and answers a question about the picture's attributes. The Phonological Awareness subtest measures elision and blending abilities. The child says words

after being instructed to drop sounds (elision) and combines separate sounds into a word after listening to the sounds (blending). Each subtest contains item sets, which are groups of items assessing the same skill. The Print Knowledge and Phonological Awareness subtests contain several item sets, whereas the Definitional Vocabulary subtest contains one item set. The assessor administers all item sets within each subtest. All three subtests have a ceiling rule of three consecutive incorrect responses, which are applied to each item set within each subtest.

Other Languages: The precursor to the TOPEL, the Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP) (Lonigan et al. 2002; Lonigan 2002), includes a Spanish language version that has been used in some studies, although no information is available on the norming sample or equivalence with the English version of the Pre-CTOPPP.

Uses of Information: The TOPEL is used to quantify and measure change over time in literacy-related abilities. The developers note that the assessment may also be used (1) to identify children at risk of having or developing literacy-related problems and (2) to monitor progress in early literacy-related skills in response to an intervention or program.

Reliability:

1. Internal consistency reliability: Cronbach's alphas were calculated for composite and subtest scores. The coefficient for the 3-year-old composite score was .95, and the

coefficient for the Print Knowledge, Definitional Vocabulary, and Phonological Awareness subtests were .93, .94, and .86, respectively.

2. Test-retest reliability: the sample consisted of 45 3- to 5-year-olds from Mandan, North Dakota, who were primarily white and female. Test-retest reliability (with a two-week interval) of standard scores ranged from .81 to .89 on subtests and was .91 on the composite score. The authors noted but did not elaborate on a practice effect for the Phonological Awareness subtest.
3. Inter-rater reliability: two trained assessors independently scored 30 randomly selected protocols from the normative sample. Reliability coefficients using standard scores for subtests and the composite ranged from .96 to .98.

Validity:

1. Content validity: TOPEL developers noted that the Print Knowledge and Phonological Awareness subtests were based on their research over the past decade. The Definitional Vocabulary subtest contained frequently used word items from several sources, such as word frequency guides, works of literature, popular fiction and non-fiction used in schools, and early vocabulary lists and analyses. Developers described various field tests with preschool-age children, generally from Florida, using an iterative process to reduce the pool of items. Items were removed

or modified based on inconsistency of children's response patterns, item difficulty level, or low correlations between items and total scores. Developers analyzed item validity and item difficulty of finalized subtests on the full normative sample. In 3-year-olds, median item discrimination coefficients ranged from .38 to .54. Median item difficulty, which reflects the percentage of children who passed a given item, ranged from .20 to .51.

2. Concurrent validity: developers correlated scores on the three TOPEL subtests with scores on the Test of Early Reading Ability-Third Edition (TERA-3) Alphabet subtest, the TERA-3 Reading Quotient, the Expressive One-Word Picture Vocabulary Test-2000 Edition (EOWPVT), the Get Ready to Read! Screening Tool, and the Comprehensive Test of Phonological Processing (CTOPP) Elision Blending Words subtests. The sample consisted of 154 3- to 5-year-olds from Tallahassee, Florida, of whom the majority was male (60 percent) and white (89 percent). Uncorrected correlations between the TOPEL Composite Early Literacy Index and the TERA-3 Reading Quotient and the Get Ready to Read! Screening Tool were 0.63 and 0.60, respectively. The TOPEL Print Knowledge subtest scores correlated .74 with the TERA-3 Alphabet scores. The TOPEL Definitional Vocabulary subtest scores correlated .62 with the EOWPVT scores. The TOPEL Phonological Awareness subtest scores correlated .52 and .55 with the CTOPP Elision and the Blending of Words scores,

respectively. In addition, the three individual TOPEL subtests were correlated with the TERA-3 Reading Quotient and Get Ready to Read!, with uncorrected coefficients ranging from .37 to .57. Developers also provided corrected correlations to account for the effects of restricted age range.

With respect to subgroup differences, the authors examined age and Hispanic American-bilingual status in relation to TOPEL performance for the entire normative sample. Chronological age was positively related to TOPEL performance on the three subtests such that raw score means increased with age. Coefficients between subtest raw scores and age were .49, .54, and .56 for Phonological Awareness, Definitional Vocabulary, and Print Knowledge, respectively. Hispanic American-bilingual children demonstrated standard scores below the average range of 90 to 110 for Definitional Vocabulary, Phonological Awareness, and the composite score (mean = 82, 89, 84, respectively) and scores below the norming sample mean of 100 but within the normal range for Print Knowledge (mean = 92). Developers noted in the manual that below-average scores for Hispanic American-bilingual children support the assessment's validity.

Bias Analysis: Three types of analyses were conducted to examine the impact on various groups of test takers: (1) Differential Item Functioning (DIF), (2) comparison of mean

standard scores, and (3) internal consistency Cronbach's alpha coefficients. DIF analysis was conducted on the entire normative sample of 3- to 5-year-olds, and the comparison groups included gender, race (black versus non-black), and ethnic (Hispanic versus non-Hispanic) groups. Developers neither reported the groups favored in each item comparison nor removed any items based on DIF analyses, but they reported several other findings. In the DIF by gender analysis, one item in the Definitional Vocabulary subtest had a moderate effect size. In the DIF by race analysis, one item in the Print Knowledge subtest and one item in the Definitional Vocabulary subtest had moderate effect sizes. In the DIF by ethnicity analysis, four items in the Definitional Vocabulary subtest had moderate or large effect sizes. The mean standard scores by gender, race, and ethnicity were average, with standard scores ranging from 92 to 105 (within the normal range of 90 to 110), and developers concluded that the TOPEL is nonbiased in regard to gender, race, or ethnicity. Cronbach's alphas were calculated for the composite and each subtest score by subgroup (male, female, white, black, and Hispanic) and ranged from .85 to .97.

Methods of Scoring: Assessors code each correct response as "1" and each incorrect response as "0." Raw scores reflect the total correct responses in all item sets to the last item in the ceiling. A total composite score (Composite Early Literacy Index) and subtest scores are computed. The manual includes appendices with conversions of raw scores into standard scores (mean of 100

and standard deviation of 15) and percentile ranks.

Interpretability: The manual includes extensive instructions for interpreting below-average, average, and above-average standard scores for the subtests and the composite measure as well as general information on what standard scores mean. Developers indicate that standard scores provide the clearest indication of a child's performance on the TOPEL. The manual briefly discusses interpretations of raw scores and percentile ranks.

Training Support: The manual provides information on the basic administration of the assessment. Developers recommend self-training by reading the manual and completing five test administrations to prepare for administration.

Adaptations/Special Instructions for Individuals with Disabilities: No information available.

Report Preparation Support: No information available.

References:

Lonigan, Christopher J., Richard K. Wagner, Joseph K. Torgesen, and Carol A. Rashotte. *Preschool Comprehensive Test of Phonological and Print Processing (Pre-CTOPPP)*. Florida State University, Tallahassee, FL: Department of Psychology, 2002.

Lonigan, Christopher J., Richard K. Wagner, Joseph K. Torgesen, and Carol A. Rashotte.

TOPEL: Test of Preschool Early Literacy. Austin, TX: PRO-ED, 2007.

Lonigan, Christopher J., Richard K. Wagner, Joseph K. Torgesen, and Carol A. Rashotte.

TOPEL: Test of Preschool Early Literacy Examiner's Manual. Austin, TX: PRO-ED, 2007.

Lonigan, C. J. "Pre-CTOPP Subtest Statistics for Preschool Comprehensive Test of Phonological and Print Processing by Age Group." Available at [<http://www.psy.fsu.edu/~lonigan/data.pdf>], 2002.

Madle, Ronald A. "Review of the Test of Preschool Early Literacy." In *The Eighteenth Mental Measurements Yearbook*, edited by Robert A. Spies, Janet F. Carlson, and Kurt F. Geisinger. Lincoln, NE: The Buros Institute of Mental Measurements, 2010.

Owens, Gretchen, and Claire Lenz. "Review of the Test of Preschool Early Literacy." In *The Eighteenth Mental Measurements Yearbook*, edited by Robert A. Spies, Janet F. Carlson, and Kurt F. Geisinger. Lincoln, NE: The Buros Institute of Mental Measurements, 2010.

TODDLER ATTACHMENT SORT-45 (TAS-45), 2004 (2011 Update)

Authors:

John Kirkland, David Bimler, Andrew Drawneek, Margaret McKim, and Axel Scholmerich

Publisher:¹

NCAST Programs
(206) 543-8528
<http://www.ncast.org>

Initial Material Cost:

A computer-based scoring software program called the Toddler Observation Training System (TOTS) is available for free at <http://www.suchandsuch.biz/tots/>, although users are charged a fee for instructions and an access code.

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

12 through 36 months old.

Personnel, Training, Administration, and Scoring

Requirements: Administered by highly trained field observers.

Susan Spieker and colleagues from the University of Washington provide TAS-45 training, which lasts about two days, plus an additional day and one-half to prepare for the training and to evaluate the group's coding reliability. Training costs \$1,000 to \$1,500 per day plus travel costs, although the University of Washington should be contacted for exact rates. Trained and certified individuals may train others in a one-day training session. One study documented that training could be as few as 5 hours by using a combined format of computer-based and in-person group training (Andreasson et al. 2007).

Observations last 60 to 90 minutes, and ratings conducted after the observation take about 10 to 20 minutes.

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 2 (most under .65 for test-retest) and 3 (most over .65 for inter-rater reliability)

Validity: 2 (most under .5 for concurrent)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 3 (administered by a highly trained individual and scored by a computer-based software program)

¹ The instrument and scoring system will not be made available until 2012.

Description: The Toddler Attachment Sort-45 (TAS-45) is an observational measure designed to assess the quality of toddlers' attachment to parents—the emotional bond that is the basis for a child's sense of security. A trained observer completes the TAS-45 ratings privately outside the home after observing the parent and toddler in a home-based setting.

The TAS-45 includes 45 descriptive statements that adhere to one of nine "hotspots," or categories of behavior. Each of the 45 statements among the nine hotspots appears twice, for a total of 90 statements. The 90 statements are clustered into 30 sets of 3, for a total of 30 items.

The nine hotspots are (1) Warm and Cuddly, (2) Cooperative, (3) Sociable, (4) Independent, (5) Attention Seeking, (6) Upset by Separation, (7) Avoids Others, (8) Demanding or Angry, and (9) Moody, Unsure, Unusual. The hotspots are related to patterns of secure and insecure parent-child attachment relationships that have been identified in the attachment literature: avoidant (A), secure (B), ambivalent (C), and disorganized (D) (Andreassen, Fletcher, and Park 2007). The TAS-45 removed statements and shortened the administration time of the Attachment Q Sort (AQS) for use in the birth cohort of the Early Childhood Longitudinal Study (Andreassen et al. 2007).²

Other Languages: None.

Uses of Information: Researchers may use TAS-45 data to classify children according to avoidant, secure, ambivalent, and disorganized styles of attachment and to examine the quality of parent-child relationships.

Reliability:

1. Internal consistency reliability: no information available.
2. Test-retest reliability: Spieker et al. (2011) analyzed a sample of 23 racially diverse mother-toddler dyads enrolled in home-based

² The AQS rating process involves piling 90 cards, each with one descriptive statement, in two waves for a “nine-pile sort.” The first wave involves placing cards in three piles (statements that do, do not, and maybe apply), and the second wave involves separating each of the three piles into three more piles ranging from highly characteristic to highly uncharacteristic of the toddler.

Early Head Start programs. The assessment interval ranged from six to eight months. Correlation coefficients for the Security Factor and Dependence Factor scores were .54 and .58, respectively. The disorganized (D) attachment category was correlated at .84.

3. Inter-rater reliability: a TAS-45 co-author and another observer completed the TAS-45 for 11 home visits to measure inter-rater reliability (Spieker et al. 2011). Pearson correlations for the Security Factor and Dependence Factor scores were .83 and .92, respectively. Pearson correlations for the nine hotspots ranged from .57 (Demanding or Angry) to .91 (Sociable, Independent, and Avoids Others). Observers reached 100 percent agreement on ABCD Classification scores. Items 14 and 30 have been reported as difficult to code (M. Oxford, personal communication, February 8, 2010).

Validity:

1. Content validity: Kirkland and colleagues (2004) collected AQS data sets from researchers in several countries and subjected the data to multidimensional scaling and facet cluster analysis to map the items onto eight dimensions. The resulting TAS-39 strongly associated 39 statements to the eight dimensions (i.e., hotspots). Field tests resulted in replacing statements to improve wording and ease of observation. On a separate set of 42 statements characterizing disorganized attachment, the authors used multidimensional scaling and facet cluster analysis to identify 12 strongly associated

statements, of which 6 were removed because they were less conducive to field observations. The remaining 6 statements were added to the TAS-39 to yield the TAS-45.

2. **Concurrent validity:** Spieker et al. (2011) found significant correlations between TAS-45 D Classification scores and Secure Factor scores and other measures. TAS-45 D Classification scores were correlated with the 2001 Brief Infant-Toddler Social and Emotional Assessment (BITSEA) Problem and Competence domain scores at .59 and -.55, respectively, and with the Preschool Language Scale, Fourth Edition (PLS-4) Expressive and Receptive Communication scale scores at -.27 and -.43, respectively. D Classification scores also correlated with a measure of parent-child communication developed by Spieker et al. (2011) to measure verbal and non-verbal communication attempts and reciprocity by parents and toddlers. The correlation between D Classification scores and toddlers' non-response to communication initiated by mothers was .30. TAS-45 Security Factor scores were correlated with BITSEA Problem and Competence domain scores at -.64 and .61, respectively, and with PLS-4 Expressive and Receptive Communication scale scores at .32 and .41, respectively. The coefficient between TAS-45 Security Factor scores and parent-child communicative reciprocity scores was .35.

3. **Construct validity:** descriptive data from ECLS-B suggest that the TAS-45 yields similar patterns of attachment across attachment classifications-avoidant (A), secure (B), ambivalent (C), and disorganized (D)-as that found in the attachment literature (Andreasson et al. 2007). For example, the TAS-45 classified 16 percent of toddlers as A, and the attachment literature suggests approximately 15 to 20 percent of children may be classified as A.

Methods of Scoring: For each item (i.e., set of three statements), the observer indicates which statement is "most true" of the toddler and which statement is "least true." This forced ranking among the three statements is termed a trilemma. Rankings are entered into the Toddler Observation Training System (TOTS) computer-based software program, which generates several types of scores as well as attachment and hotspot score profiles.

The TOTS program generates four types of scores: Traditional (or ABCD) Classification scores, Confidence scores, Security Factor scores, and Dependency Factor scores. Traditional classification scores are the child's TAS-45 profile mapped onto to the ABCD attachment classification system. Confidence scores are an indicator of the distance between the TAS-45's profiles and the traditional classification system. That is, the Confidence score is an indicator of how well the TAS-45 profile maps onto the traditional classification system. Confidence scores range from 0 to 1 with higher scores

indicative of greater confidence in the TAS-45 profile. The TAS-45's Security Factor and Dependency Factor scores are generated by using the same published criterion sorts used by the AQS.³ The relationship between Security and Dependency factor scores may be indicative of the traditional ABCD classification system scores. For example, a low Security Factor score and a high Dependency Factor score may suggest traditional Attachment C (Andreasson et al. 2007). Security Factor scores and ABCD Classification scores are recommended for researchers examining toddler attachment (Andreasson et al. 2007).

Interpretability: Attachment and hotspot score profiles provide snapshots of children's attachment characteristics during the observation period. The hotspot profiles represent the child's scores on the nine TAS-45 hotspots and indicate the degree to which the child displayed each of the nine behaviors during the observation period. Hotspot profile scores are used to show toddlers' overall attachment security and may be plotted at different ages. The attachment profiles align scores on the hotspot profiles with the traditional ABCD attachment classification system and generate the child's profile on the ABCD system.

Training Support: Published studies have documented training methods on the TAS-45 (Andreasson et al. 2007; Spieker et al. 2011). Formats used in the studies included computer-based training modules, in-person group trainings, and practice during pilot study home

observations. Three computer-based training modules each conclude with a quiz that trainees need to pass with at least an 80 percent score. Modules cover an introduction to the TAS-45 and sorting, a review of items and their applications, and a review of video clips of children with different attachment characteristics. For the ECLS-B study, Kirkland calculated percentage agreement scores with his scores to ensure that they were reliable at 80 percent agreement (Andreasson et al. 2007). Spieker and colleagues (2011) compared pilot study ratings item by item and estimated inter-coder agreement. More recently, the University of Washington has provided group training by using a non-computer-based format (S. Monahan, personal communication, June 8, 2011). This training involved self-administered practice exercises, in-person training that covered a review of the foundations of attachment theory and the nine TAS-45 hotspots, practice using the instrument with video clips of children, and two webinars after the training for review and practice. The training concluded with a reliability assessment in which trainees had to score at least 85 percent agreement with the trainer.

Report Preparation Support: The TOTS program generates attachment and hotspot score profile reports for individual toddlers at different ages. Condon and Spieker (2008) suggest that discussions with parents about TAS-45 findings should focus on understanding children's attachment strategies as part of families' effort to

³ Waters et al. (1995) provide the criterion sorts for generating the Security and Dependency Factor scores.

create a nurturing relationship environment for their children.

References:

- Andreassen, Carol, Philip Fletcher, and Jennifer Park. "Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) Psychometric Report for the 2-Year Data Collection." Report No. NCES 2007-084. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, 2007.
- Condon, Marie-Celeste, and Susan Spieker. "It's More than a Measure: Reflections on a University-Early Head Start Partnership." *Infants & Young Children*, vol. 21, no. 1, 2008, pp. 70-81.
- Kirkland, John, David Bimler, Andrea Drawneek, Margaret McKim, and Axel Scholmerich. "An Alternative Approach for the Analyses and Interpretation of Attachment Sort Items." *Early Childhood Development and Care*, vol. 174, no. 7-8, 2004, pp. 701-719.
- Spieker, Susan, Elizabeth M. Nelson, and Marie-Celeste Condon. "Validity of the TAS-45 Measure of Toddler-Parent Attachment: Preliminary Evidence from Early Head Start Families." *Attachment and Human Development*, vol. 13, no. 1, 2011, pp. 69-90.
- Waters, Everett, and Kathleen E. Deane. "Defining and Assessing Individual Differences in Attachment Relationships: Q-Methodology and the Organization of Behavior in Infancy and Early Childhood." *Monographs of the Society for Research in Child Development*, vol. 50, no. 1-2, 1985, pp. 41-65.
- Waters, Everett, Brian E. Vaughn, German Posada, and Kiyomi Kondo-Ikemura. "Caregiving, Cultural, and Cognitive Perspectives on Secure-Base Behavior and Working Models: New Growing Points of Attachment Research." *Monographs of the Society for Research in Child Development*, vol. 60 (Serial No. 244), no. 2-3, 1995.

VINELAND ADAPTIVE BEHAVIOR SCALES, SECOND EDITION (VINELAND-II), 2005 (2011 Update)

Authors:

Sara S. Sparrow, Domenic V. Cicchetti, David A. Balla

Publisher:

NCS Pearson, Inc.

(800) 627-7271

<http://www.pearsonassessments.com>

Instrument:

<http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=Vineland-II&Mode=summary>

Initial Material Cost:

Vineland-II Survey Forms Starter Set (includes 10 Survey Interview Forms, 10 Parent/Caregiver Rating Forms, 10 Survey Interview Reports to Parents, 10 Survey Forms Reports to Caregivers, and 1 Survey Forms Manual): \$162.30

Vineland-II Survey Forms ASSIST Mac/Win (includes scoring and reporting software for the Survey Interview Form and the Parent/Caregiver Rating Form; for Windows and Macintosh): \$269.70

Representativeness of Norming Sample:

The nationally representative norming sample includes 3,695 individuals age birth through 90 years. Random sampling methods were used to select a sample that closely matched national norms in the *Current Population Survey, March 2001* (including age, sex, geographic region, parent's education, race/ethnic group, and community size).

Languages:

English, Spanish

Type of Assessment:

Parent/caregiver report

Age Range and Administration Interval:

Birth through 90 years

Personnel, Training, Administration, and Scoring Requirements:

Interviewers should have graduate-level education in psychology or social work as well as in individual assessment and test interpretation. Interviewers must read and study the Vineland-II Survey Forms Manual before administration (practice sessions recommended).

Approximately 20 to 60 minutes to administer the Survey Interview Form and 15 to 30 minutes to score.

A Vineland-II Training CD is available from the publisher for \$110.10.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (mostly .65 or higher)

Validity: 3 (mostly .5 or higher for concurrent and construct)

Norming Sample Characteristics: 3 (normed within past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Vineland-II Adaptive Behavior Scales (Vineland-II) measure personal and social skills from birth through age 90 and were designed to address special needs populations, including individuals with mental retardation, autism, and attention-deficit/hyperactivity disorder (ADHD). There are

383 items in the Adaptive Behavior Composite, and 50 additional items on the Maladaptive Behavior Index. The scale may be administered by using the Survey Interview Form (a semistructured interview) or the Parent/Caregiver Rating Form (a rating scale). Using the semistructured interview format, the interviewer

asks general open-ended questions to the parent or caregiver relating to the child's activities and behavior to ascertain key developmental milestones. The format requires the interviewer to design his or her own open-ended questions; therefore, it is critical that the interviewer has a thorough understanding of the test items as well as experience in conducting such interviews. Using the rating scale format, a parent or caregiver familiar with the behavior of the infant or child undergoing assessment rates a checklist of the individual's skills. An Expanded Interview Form for infant and children birth through 5 years provides an in-depth alternative to the Survey Interview Form, with additional items. A Teacher Rating Form assesses adaptive behavior for children in day care, preschool, or school. The Vineland-II updates the previous version of the Vineland Adaptive Behavior Scales (1984) with new norms, expanded age range, updated content, increased coverage of early childhood adaptive behavior, streamlined administration, and the Parent/Caregiver Rating Form.

The Adaptive Behavior Composite comprises four domains: Communication, Daily Living Skills, Socialization, and Motor Skills. Each domain consists of subdomains. The Communication domain measures receptive, expressive, and written communication; the Daily Living Skills domain assesses personal, domestic, and community skills; the Socialization domain measures interpersonal relationships, play and leisure time, and coping skills; and the Motor Skills domain measures gross and fine motor skills. As an optional fifth domain for individuals

age 3 through 90 years, the Maladaptive Behavior domain, includes subdomains for internalizing, externalizing, and other types of undesirable behavior.

Other Languages: The Vineland-II includes a record booklet with a Spanish translation of the Survey Interview Form, which was used with Spanish-speaking respondents during the national standardization. Detail on the reliability, validity, and English language equivalence is not available.

Uses of Information: The Vineland-II assesses an individual's daily functioning. It may be used in educational and clinical diagnostic evaluations of developmental delays, in developmental evaluations of young children, for progress monitoring, and for program planning. It may also be used in research projects to determine the effects of various treatments or clinical interventions on adaptive functioning.

Reliability:

1. Internal consistency reliability: split-half reliability estimates using Spearman-Brown correlation coefficients for infants and toddlers birth through 36 months ranged from .79 (Daily Living Skills) to .95 (Socialization) and from .95 to .98 among age groups for the Adaptive Behavior Composite. Coefficients for children age 4 and 5 years ranged from .83 (Motor Skills) to .93 (Socialization) and to .97 for the Adaptive Behavior Composite.¹ Cronbach's alpha for

¹ Split-half reliability coefficients are corrected for half-test length by the Spearman-Brown formula.

children age 3 through 5 years on the Maladaptive Behavior Index was .88.

2. Test-retest reliability: for children birth through 2 years, adaptive behavior domain ICCs ranged from .86 (Motor Skills) to .95 (Communication); the Adaptive Behavior Composite ICC was .96 (with administration intervals ranging from 14 to 30 days [mean = 18 days]). For children 3 through 6 years, adaptive behavior domain ICCs ranged from .88 (Socialization) to .90 (Communication and Daily Living Skills); the Adaptive Behavior Composite ICC was .94 (with administration intervals ranging from 14 to 30 days [mean = 20.6 days]). For ages 3 through 5 years, the Maladaptive Behavior Index ICC was .89 (with administration intervals ranging from 14 to 30 days [mean = 20.5 days]).
3. Inter-rater reliability: for a group of 152 respondents (primarily parents) in which two respondents each completed the Parent/Caregiver Form for the same individuals, ICCs for children birth through 6 years ranged from .61 (Daily Living Skills) to .82 (Motor Skills); the Adaptive Behavior Composite ICC was .80. The ICC on the Maladaptive Behavior Index for children age 3 through 11 years was .81. ICCs were calculated among interviewer pairs who completed the Survey Interview Form for children birth through 6 years. Coefficients ranged from .58 (Socialization) to .82 (Daily Living Skills); the coefficient for the Adaptive Behavior Composite was .79. The ICC on the

Maladaptive Behavior Index for children age 3 through 11 was .83.

Validity:

1. Content validity: confirmatory factor analysis evaluated the fit of the model. For children 6 years and younger, the Adaptive Behavior Composite is based on a four-factor solution that had a comparative fit index of .96.
2. Concurrent validity: the Vineland-II was compared with the Vineland Adaptive Behavior Scales (1984); the Adaptive Behavior Assessment System, Second Edition (ABAS-II); and the Behavior Assessment System for Children, Second Edition (BASC-2). Between the Vineland-II and Vineland Adaptive Behavior Scales (1984) for children birth through 2 years, coefficients ranged from .65 to .91 across the adaptive behavior domains; the coefficient was .82 for the Adaptive Behavior Composite. The coefficients ranged from .85 to .94 across the behavior domains for children 3 through 6 years; the coefficient was .91 on the Adaptive Behavior Composite. The coefficient between the composite scores on the Vineland-II and ABAS-II for children birth through 5 years was .63.

The coefficient between the Vineland-II and BASC-2 for children 3 through 5 years was .45 for the adaptive behavior composite scores, ranging from .35 to .59 across domains. The coefficient between the BASC-2 and Maladaptive Behavior Index was .53, ranging from .29 to .49 across domains.

In terms of capturing subgroup differences, the Vineland-II was used as a measure of adaptive functioning to diagnose autism in children age 3 through 16 years in groups with and without autism. Differences of more than two standard deviations between the Adaptive Behavior Composite and domain scores between groups were observed.

3. Construct validity: adaptive behavior domain intercorrelation coefficients (ICCs) ranged from .56 (Communication and Motor Skills) to .65 (Communication and Socialization) for children 0 to 2 years old and from .61 (Communication and Motor Skills) to .73 (Daily Living Skills and Socialization) for children 3 to 6 years old. ICCs between adaptive behavior domains and the Adaptive Behavior Composite ranged from .82 to .87 for 0- to 2-year-olds and from .83 to .88 for 3- to 6-year-olds. The identification of the adaptive behaviors and skills measured by the Vineland-II and the development of test content are closely linked to the instrument's theoretical structure. The structure of adaptive behavior functioning for the Vineland-II is supported by the Vineland Adaptive Behavioral Scales, the American Association of Mental Retardation (2002), the American Psychological Association (1996), and the National Academy of Sciences (2002).

Bias Analysis: Using differential item functioning (DIF), the authors tested for measurement bias by sex, parental education, and ethnicity (adjusted for socioeconomic status and

sex) at both the item and scale levels. Children with mothers who did not complete high school showed fewer adaptive behaviors on average than children whose mothers had at least a high school diploma, although the differences were described as small, with no other differences observed.

Methods of Scoring: Items within the adaptive behavior subdomains are scored according to whether the item activity occurs usually (2), sometimes or partially (1), never (0), or don't know (DK). The assessor must first identify basal and ceiling items as described in the Survey Forms Manual. The raw score is equal to the sum of 1 and 2 responses, plus the number of items before the basal item multiplied by two. If there are more than two DK and/or missing responses, then a subdomain score may not be obtained. The Maladaptive Behavior Index items are scored according to whether the item behavior occurs often (2), sometimes (1), or never (0) for the Problem Behaviors Parts 1 and 2. For the Problem Behaviors Part 2, items are also scored according to whether the behavior is severe (S) or moderate (M). Subdomain raw scores are the sum of items with a 1 or 2 response. Subdomain raw scores are added to obtain the Maladaptive Behavior Index raw score. Raw scores, v-scale scores (mean of 15 and standard deviation of 3), standard scores (mean of 100 and standard deviation of 15), percentile ranks, adaptive levels (low, moderately low, adequate, moderately high, and high), age equivalents, and stanines are available in the manual. Scores at high levels of adaptive functioning have poorer precision

(higher standard errors of measurement) than at lower levels (Sattler and Hoge 2006), but Stein (2010) comments that given the Vineland-II's focus on the assessment of clinical syndromes such as mental retardation and autism, the precision issue should not be a concern.

Interpretability: The Vineland-II should be interpreted by professionals with graduate degrees and specialized training and experience in administering and interpreting assessments. The Survey Forms Manual contains a chapter with steps and examples on how to interpret children's scores to identify children's strengths and weaknesses and facilitate communications with parents and caregivers.

Training Support: A Vineland-II Training CD is available from the publisher.

Adaptations/Special Instructions for Individuals with Disabilities: The Vineland-II was designed to address special needs populations, including individuals with mental retardation, autism, and ADHD. Administration instructions do not vary across groups.

Report Preparation Support: The Score Report (in both the Survey Interview and Parent/Caregiver Rating forms) includes a Score Summary page, Score Profile page, and Pairwise Comparisons page. On the Score Summary page, the assessor records subdomain, domain, Adaptive Behavior Composite, and Maladaptive Behavior Index scores. On the Score Profile page, the assessor records domain standard scores or subdomain v-scale scores and their respective

confidence intervals. The Pairwise Comparison page allows for statistical analyses of the score differences between pairs of subdomains and pairs of domains. The Survey Forms Manual includes steps to complete the comparisons.

The Vineland-II ASSIST software calculates derived scores and produces reports, including the adaptive behavior score summary, domain and subdomain analyses, graphical profiles, narrative reports, and a caregiver letter.

References:

Pearson. *Vineland-II ASSIST Scoring and Reporting Software*. 2005.

Pearson. *Vineland-II Training CD*. 2005

Sattler, Jerome M. and Robert D. Hoge. *Assessment of Children: Behavioral, Clinical, and Social Foundations*. San Diego, CA: Jerome M. Sattler Publisher Inc., 2006.

Sparrow, Sara S., David A. Balla, and Domenic V. Cicchetti. *Vineland Adaptive Behavior Scales*. Bloomington, MN: Pearson Assessments, 1984.

Sparrow, Sara S., Domenic V. Cicchetti, and David A. Balla. *Vineland Adaptive Behavior Scales, Second Edition*. Minneapolis, MN: Pearson Assessments, 2005.

Sparrow, Sara S., Domenic V. Cicchetti, and David A. Balla. *Vineland-II Adaptive Behavior Scales Record Booklet*. Circle Pines, MN: AGS Publishing, 2005.

Stein, Stephanie. "Review of the Vineland Adaptive Behavior Scales, Second Edition

(Vineland-II).” In *The Eighteenth Mental Measurements Yearbook*, edited by Robert A. Spies, Janet F. Carlson, and Kurt F. Geisinger. Lincoln, NE: The Buros Institute of Mental Measurements, 2010.

Widaman, Keith F. “Review of the Vineland Adaptive Behavior Scales, Second Edition

(Vineland-II).” In *The Eighteenth Mental Measurements Yearbook*, edited by Robert A. Spies, Janet F. Carlson, and Kurt F. Geisinger. Lincoln, NE: The Buros Institute of Mental Measurements, 2010.

VINELAND SOCIAL-EMOTIONAL EARLY CHILDHOOD SCALES/VINELAND SEEC, 1998

Authors:

Sara S. Sparrow, David A. Balla, Domenic V. Cicchetti

Publisher:

AGS (American Guidance Service, Inc.)
(800) 328-2560
www.agsnet.com

Initial Material Cost:

Vineland SEEC Kit: \$55 (includes Manual and 25 Record Forms) Vineland SEEC ASSIST CD Kit: \$200

Representativeness of Norming Sample:

Standardization norms based on the normative data used to develop the Vineland Adaptive Behavior Scales. The sample included 1,200 children from birth to 5 years, 11 months selected to closely match the national norms in the 1980 U.S. census (including age, gender, geographic region, parent education, race/ethnic group, and community size).

Languages:

English (reports to parents also available in Spanish)

Type of Assessment:

Interviewer-assisted parent report

Age Range and Administration Interval:

0 through 5 years, 11 months

Personnel, Training, Administration, and Scoring

Requirements: Manual suggests that interviewers should have graduate education in early childhood development and training in interview techniques and experience in administering, scoring, and interpreting the Vineland SEEC Scales in practice sessions. Requires approximately 25 minutes to administer the assessment, and 10 to 15 minutes for scoring.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: Internal Consistency and Test-Retest: 3 (.65 or higher)

Validity: 3 (.5 or higher for consistency)

Norming Sample Characteristics: 2 (older than 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Vineland SEEC Scales assess the social-emotional functioning of children from birth through 5 years, 11 months. Three scales, which combine into a Social-Emotional Composite, are used to evaluate a child's ability to pay attention, understand emotional expression, cooperate with others, construct and observe relationships, and develop self-regulation behaviors. The three scales are: Interpersonal Relationships (44 items), Play and Leisure Time (44 items), and Coping Skills (34 Items). This assessment is administered as a semi-structured interview with the child's parent or caregiver, in which the interviewer asks general open-ended

questions relating to the child's activities and behavior (these questions are designed by the interviewer) to ascertain key developmental milestones. Since this requires that the interviewer design his or her own open-ended questions, it is critical that the interviewer has a thorough understanding of the test items and experience in conducting this type of interview.

Uses of Information: The Vineland SEEC can be used in educational and clinical settings to help identify developmental delays, plan and select appropriate activities for young children, and monitor developmental progress. The SEEC can also be used in research projects to determine the

effects of various treatments or clinical interventions on young children's social-emotional functioning. The manual suggests that in order to obtain a more comprehensive description of children's development, it is useful to use the Vineland SEEC measures along with other measures of children's physical, cognitive, language and adaptive skills.

Reliability: The reliability statistics for children between 6 and 36 months indicate (1) Internal consistency (Spearman-Brown correlations): Interpersonal Relationships: .82 to .92; Play and Leisure Time: .72 to .96; Coping Skills: .87; and Composite: .89 to .97 (2) Test-retest reliability (interval ranged from 2 to 4 weeks and averaged 17 days): Interpersonal Relationships: .73; Play and Leisure Time: .74; Coping Skills: .54; and Composite: .77. (3) Inter-rater reliability (intervals ranged from 1 to 14 days and averaged 8 days): .47 to .60.

Validity: No validity studies are included in the manual for the Vineland SEEC. The manual states that since the item content and scale structure of the Vineland SEEC replicate the socialization domain of the Vineland Adaptive Behavior Scale (ABS), its results are generalizable to the SEEC. Concurrent validity studies on the Vineland ABS indicate correlations with similar established tools range from .51 to .65.

Method of Scoring: Items in the SEEC are scored based on how often the child performs the specific activity. Options are: "usually performs," "sometimes or partially performs," "never performs," "no opportunity for the child to

perform," and "don't know if the child performs." The Appendix of the SEEC contains scoring criteria for each of the scales. In addition, there is a software program—the Vineland SEEC ASSIST (Automated System for Scoring and Interpreting Standardized Tests)—available to help score and interpret the Vineland SEEC. The user can input raw scores or item scores to obtain a derived score and an interpretive report.

Interpretability: The manual contains a chapter on interpretation of scores as well as case studies on how to interpret a child's performance on the Vineland SEEC Scale. The manual includes age-based standard scores for 1-month intervals from birth through age 2, and at 2-month intervals from age 2 to age 5 years, 11 months. It is suggested that the Vineland SEEC Scale be implemented by professionals with graduate degrees and specialized training and experience in administering and interpreting early childhood assessments.

Training Support: Dr. Sara S. Sparrow (one of the authors), is available to conduct professional workshops on effectively administering and interpreting the Vineland, but these must be specifically arranged with the publisher.

Adaptations/Special Instructions for Individuals with Disabilities: Individuals classified in special education categories were included in the standardization sample for the Vineland SEEC Scales. The manual also directs the interviewer to derive an adjusted age for those children who were born prematurely. There is

limited information in the manual regarding adaptation for children with disabilities—the manual suggests that the interviewer should follow scoring guidelines and criteria while taking into account the specific disability (for example, a child may use sign language to perform some activity described in the items on the scale).

Report Preparation Support: The Program Planning Report is meant to be used as an outline for recommended education and/or treatment plans. The Vineland SEEC ASSIST software produces reports on: personal information summary, score profile, score narrative, program

planning profile, letter to parents (in English or Spanish), and recommended activities.

References:

Sparrow, Sara S., David A. Balla, and Domenic V. Cicchetti. Vineland Social-Emotional Early Childhood Scales/Vineland SEEC. Circle Pines, MN: American Guidance Service, Inc., 1998.

Sparrow, Sara S., David A. Balla, and Domenic V. Cicchetti. Vineland SEEC ASSIST (Automated System for Scoring and Interpreting Standardized Tests). Circle Pines, MN: American Guidance Service, Inc., 2001.

WOODCOCK-JOHNSON III NORMATIVE UPDATE (WJ III NU), 2007 (2011 Update)

Authors:

Richard W. Woodcock, Kevin S. McGrew, and Nancy Mather

Publisher:

Riverside Publishing
(800) 323-9540
<http://www.riverpub.com>

Instrument:

<http://www.riverpub.com/products/wjIIIComplete/>

Initial Material Cost:

WJ III NU Complete Battery (includes Cognitive Standard and Extended Test Books, Examiner's Manual, Examiner's Training Workbook, Audio Recording, 25 Test Records and 25 Response Booklets, 5 Brief Intellectual Ability Test Records, Achievement Form A Standard and Extended Test Books, Examiner's Manual, Examiner's Training Workbook, Audio Recording, 25 Test Records and 25 Response Booklets, WJ III NU Compuscore and Profiles Program (Windows and Macintosh), NU Technical Manual, and Scoring Guides): \$1,322

WJ III NU Achievement Battery kit: \$590

WJ III NU Cognitive Abilities Battery kit: \$887

WJ III NU Tests of Achievement, Form C/Brief Battery: \$379

Representativeness of Norming Sample:

The WJ III NU updates the normative data for the WJ-III and Bateria III cognitive and achievement batteries. The updated norms are based on 8,782 of the 8,818 individuals in the original WJ III sample. The preschool-age sample (children 2 to 5 years) included 1,153 children. The authors weighted individual subjects to obtain a sample representative of the 2005 census. The original sample came from 100 geographically diverse U.S. communities. The developers stratified the sample by region, community size, gender, race, Hispanic/non-Hispanic background, foreign-/native-born, and school type.

Languages:

English, Spanish

Type of Assessment:

Direct child assessment

Age Range and Administration Interval:

For 10 WJ-III Cognitive and 10 WJ-III Achievement tests, age 2 to adult; for 19 WJ-III Cognitive and 13 WJ-III Achievement tests, age 4 to adult; for the remaining tests, varies from school-age to adult. Achievement tests come in two equivalent forms to allow repeat administrations.

Personnel, Training, Administration, and Scoring Requirements:

Administered by a highly trained program staff member. Assessors must provide evidence that they meet the highest standards required for administering educational and psychological tests.

Each test takes approximately 5 to 10 minutes, with the Cognitive standard battery requiring 45 to 50 minutes and the Achievement requiring 60 to 70 minutes. A computer scoring program is required to generate scores.

The publisher makes available training videos (\$43) and workbooks (\$14) and offers national and regional group training sessions as well as individual training sessions. Private training costs an average of \$800 to \$1,200 per day, not including the trainer's travel expenses.

Summary:

Initial Material Cost: 3 (>\$200)

Reliability: 3 (mostly .65 or higher)

Validity: 3 (all .5 or higher for preschool-age concurrent; mostly .4 or higher for preschool-/early elementary-age predictive)

Norming Sample Characteristics: 3 (normed within the past 15 years, nationally representative)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Woodcock-Johnson III Normative Update (WJ III NU) comprises

updated norms and norming procedures for two co-normed assessment batteries, the Woodcock-

Johnson III Tests of Cognitive Abilities (WJ III COG) and the Woodcock-Johnson III Tests of Achievement (WJ III ACH). These instruments provide a comprehensive set of norm-referenced tests for measuring intellectual abilities and academic achievement in individuals age 2 years through adulthood. The administration materials remain the same as for the previous WJ III versions. The WJ III NU updates are incorporated into the Technical Manual and WJ III NU computer scoring program. The updated Technical Manual contains new validity information as well as new methods of analyzing intra-individual variation in cognitive and achievement performance. The WJ III NU Compuscore and Profiles Program includes new parent report and summary report options that present findings in terms of standard or proficiency scores.

The WJ III COG consists of a standard battery of 10 tests, an extended battery of 10 tests (to provide in-depth assessment of different types of abilities), and 11 supplementary diagnostic tests (to pinpoint further any specific areas of weakness or strength). Five of the 10 standard COG tests may be used with children as young as 2 years of age and include (1) Verbal Comprehension, (2) Spatial Relations, (3) Sound Blending, (4) Concept Formation, and (5) Incomplete Words. Five of the 10 extended-battery COG tests may be used with children as young as 2 years of age and include (1) General Information, (2) Retrieval Fluency, (3) Picture Recognition, (4) Memory for Words, and (5) Rapid Picture Naming.

Standard- and extended-battery COG tests may be grouped to yield three overall categories of cluster scores: (1) Cognitive Performance clusters (Verbal Ability, Thinking Ability, and Cognitive Efficiency); (2) Cattell-Horn-Carroll (CHC) Factor Clusters (Comprehension-Knowledge, Long-Term Retrieval, Visual-Spatial Thinking, Auditory Processing, Fluid Reasoning, Processing Speed, and Short-Term Memory); and (3) Clinical Clusters (Phonemic Awareness, Working Memory, Broad Attention, Cognitive Fluency, Executive Processes, Delayed Recall, and Knowledge). Given the limited number of tests used with children as young as 2 years of age, some cluster scores are not applicable to toddlers.

The WJ III ACH consists of a standard battery of 12 tests and an extended battery of 10 tests (to provide in-depth assessment of an achievement area). Six of the 12 standard-battery ACH tests may be used with children as young as 2 years of age and include (1) Letter-Word Identification, (2) Story Recall, (3) Understanding Directions, (4) Spelling, (5) Passage Comprehension, and (6) Applied Problems. Four of the 10 extended-battery ACH tests may be used with children as young as 2 years of age and include (1) Picture Vocabulary, (2) Oral Comprehension, (3) Quantitative Concepts, and (4) Academic Knowledge.

The WJ III NU also introduces the WJ III Tests of Achievement Form C/Brief Battery (WJ III Form C/Brief Battery). It offers abbreviated achievement testing and scoring options for three achievement areas (Brief Reading, Brief Math, and

Brief Writing). Brief Achievement and Brief Reading may be used on children as young as 2 years of age.

Across all batteries, the assessor may tailor the administration by selecting the tests that best tap the abilities and skills of interest for a particular child. Floor and ceiling effects have been observed on some WJ III tests with children age 2 years and 5 years, 6 months (Bradley-Johnson and Durmusoglu 2005; Krasa 2007).

Other Languages: The Bateria III Woodcock-Muñoz is a Spanish adaptation of the WJ III that allows for comprehensive assessment of intellectual ability (including bilingual and low verbal ability), specific cognitive abilities, scholastic aptitude, oral language, and academic achievement in individuals from age 2 to 90 years. The manual does not specify which tests may be used for children as young as 2 years of age. All of the WJ III tests have been translated or adapted into Spanish for the Bateria III. For the Bateria III Woodcock-Muñoz: Pruebas de Habilidades Cognitivas (Bateria III COG), assessors may choose from six scales: (1) brief, (2) standard, (3) extended, (4) early development, (5) bilingual (with Diagnostic Supplement), and (6) low verbal (with Diagnostic Supplement). The Bateria III Woodcock-Muñoz: Pruebas de Aprovechamiento (Bateria III APROV) consists of five reading tests, four oral language tests, four mathematics tests, four written language tests, and four supplemental tests of academic language proficiency. The Comparative Language Index (CLI) may also be used to assess language dominance. The WJ III

NU computer scoring program provides updated norms for the Bateria III; in addition, a Spanish version of the Woodcock Interpretation and Instructional Interventions Program software is available (see Interpretability). Given that score scales are linked with those of the WJ III, individual scores on the Bateria III may be compared directly to WJ III scores. Such comparability is useful for comparing children's proficiency on assessed tasks in both Spanish and English. The computer scoring program may also compute children's cognitive-academic language proficiency (CALP).

The developers collected data from a calibration sample of 1,413 native Spanish speakers from various Spanish-speaking regions in the United States and abroad (279 were from nine U.S. states). These data were equated to WJ III norms, in that each task for each Bateria III test is scaled according to the parallel WJ III test. Using Item Response Theory (IRT) methods, the developers equated Bateria III test data to that of parallel tests on the WJ III, making the scores between the WJ III and Bateria III directly comparable. Schrank et al. (2005) reported that confirmatory factor analyses (CFA) of the Bateria III standardization data supported the measure's CHC theory-based latent factor structure (one general factor and nine broad factors) with subsamples of 6- to 13-year-olds and 14- to 19-year-olds. Patterns and magnitudes of Bateria III factor loadings demonstrated a latent factor structure similar to that of the WJ III. Bateria III internal consistency reliability coefficients for scores approximating those of the WJ III norming

sample. For 4- to 13-year-olds, coefficients ranged from .72 to .94 on cognitive battery tests and from .67 to .98 on achievement tests (Schrank et al. 2005).

Uses of Information: The WJ-III NU permits age- or grade-based norm-referenced interpretation for individual ability and achievement scores. The information may be used for diagnosis of academic strengths and weaknesses, educational programming, growth assessment, program evaluation, and research.

Reliability:

1. Internal consistency reliability: the developers calculated split-half reliability estimates for scores for all WJ III NU tests, except the timed tests and tests with multiple-point scoring systems, for which they conducted Rasch analysis procedures. For the 10 WJ III NU Tests of Cognitive Abilities that may be used with children as young as 2 years of age, reliability estimates for scores for children 2 through 3 years of age ranged from .70 to .97, with most estimates at .80 or above. For the 19 tests that may be used with children as young as 4 years of age, reliability estimates for scores for children age 4 through 5 years ranged from .78 to .98, except for Planning, which had reliability estimates ranging from .63 to .64. For the 10 WJ III NU Tests of Achievement that may be used with children as young as 2 years of age, reliability estimates for scores for 2- and 3-year-olds ranged from .56 to .98; most estimates were at .80 or above. For the 13 tests that may be used with

children as young as 4 years of age, reliability estimates for scores for 4- and 5-year-olds ranged from .61 to .99; again, most estimates were at .80 or above. The publishers recommend the use of cluster scores (i.e., groups of items from two or more tests) because such scores demonstrate consistently higher reliability.

2. Test-retest reliability: the authors reported test-retest reliability estimates for 15 WJ III cognitive and achievement tests, with intervals ranging from less than 1 year to more than 10 years. For retest intervals between 1 and 2 years, reliability estimates ranged from .57 to .91 for children age 2 to 7 years (most were above .80); for retest intervals from 3 to 10 years with the same age group, estimates ranged from .35 to .90, with tests in the Thinking Abilities cluster exhibiting lower reliability estimates than those in the Acquired Knowledge cluster. In a second study conducted with 457 students 4 to 17 years of age, researchers calculated the test-retest reliability estimates of the 17 WJ III ACH tests and 16 clusters, with a retest interval of 1 year. For students age 4 through 7 years, reliability coefficients across ages ranged from .59 to .92 for scores across the tests and from .74 to .96 for the clusters.
3. Inter-rater reliability: no information available for infant/toddler or preschool-age groups.

Validity:

1. **Content validity:** the WJ III NU's content rests on its adherence to the CHC theory of cognitive abilities for tests and clusters. Content was also designed to assess core curricular areas and areas specified in federal legislation. For the cognitive battery, experts developed test items to measure both narrow and broad abilities; each test is intended to measure a discrete narrow ability, and clusters of tests are meant to assess broad abilities. Achievement test items were developed to sample skills of oral language and academic achievement in reading, mathematics, written language, and curricular knowledge. The Technical Manual cites data demonstrating the growth and decline of cognitive and achievement abilities over the lifespan.

The developers conducted two confirmatory factor analyses with the WJ III norming sample (N = 3,900). Results indicated that, for the WJ III COG, a latent factor model with one general factor (g) and seven broad factors provided the best fit among alternative models. Data analyses on the combined cognitive and achievement batteries showed that an expanded model with one general factor and nine broad factors, plus several narrow abilities, provided the most plausible fit.
2. **Concurrent validity:** the WJ III NU Technical Manual cites positive correlations between the WJ III tests and clusters applicable to children as young as age 2 that measure similar

constructs. The developers also found positive correlations between WJ III tests and clusters and other tests measuring similar constructs. Studies conducted with preschool-age samples found correlations ranging from .67 to .76 between the WJ III General Intellectual Ability standard and extended scores with full-scale or composite scores from the Differential Ability Scales, Wechsler Preschool and Primary Scale of Intelligence-Revised, and Stanford-Binet Intelligence Scale-Fourth Edition. These studies also reported correlations ranging from .60 to .67 between the WJ III Brief Intellectual Ability score and the full-scale or composite scores from the aptitude tests mentioned above.

The WJ III NU Technical Manual presents median scores and standard deviations for selected WJ III NU clusters for the total norming sample (not broken down by age) and for 11 clinical samples comprising individuals with developmental, educational, and neuropsychological disabilities and gifted children and adolescents under age 19. Although statistical significance was not established, cluster score differences were observed across clinical groups. For example, gifted students' median cluster scores ranged from 103 to 121 versus 99 to 101 for the total norming sample.

3. **Construct validity:** the WJ III NU Technical Manual states that intercorrelations among tests and clusters measuring different abilities were lower than those between tests

measuring similar abilities. For example, correlations between the Comprehension-Knowledge tests of Verbal Comprehension, General Information, Picture Vocabulary, and Academic Knowledge with the Visual-Spatial tests of Spatial Relations and Picture Recognition ranged from .37 to .48 for 2- and 3-year-olds and from .34 to .64 for 4- and 5-year-olds. At the cluster level, intercorrelations among WJ III COG clusters typically ranged from .34 to .89 for 2- to 3-year-olds.

Bias Analysis: The developers conducted bias analyses during development of the WJ III to minimize potential bias related to gender, race, Hispanic origin, and disability status. First, experts reviewed items for potential bias and eliminated or revised all items identified as potentially biased. Next, selected items were subjected to differential item functioning (DIF) analyses conducted with the Rasch IRT model. The analyses focused on a pool of items from the WJ III COG Comprehension-Knowledge tests and the WJ III ACH Academic Knowledge test in view of the tests' strong emphases on language and achievement influences. The items assessed vocabulary, general language development, general information, and curricular and general cultural knowledge. The results indicated that only a few items differed significantly between groups; expert reviewers flagged and removed one item. The other items were retained because of the possibility of spurious findings related to the number of statistical comparisons conducted. In

addition, the developers conducted three multiple-group CFAs to examine latent factor structure invariance across groups. The latent factor structure of the WJ III was largely invariant between males and females, white and non-white students, and Hispanic and non-Hispanic students. An independent study supported the latent structural invariance of the WJ III for black and white students (Edwards and Oakland 2006).

Methods of Scoring: The Examiner's Manuals and the test easels (the flip books used for testing) summarize the general test and individual item scoring rules. The assessor indicates on the test record form whether the child passes or fails an item. The assessor computes raw scores by summing the number of correct responses and then enters the scores into the computer scoring program, which generates norm-referenced scores (computer scoring is required for the WJ III NU). Grade or age equivalents, instructional ranges, standard scores (deviation quotients), and percentile ranks may be computed for each test and cluster. Users may also compute relative proficiency indexes (RPI), which are ratios reflecting a child's performance compared to the performance of the average child of the same age or grade.

Interpretability: The Examiner's Manuals provide information about how to interpret individual test scores, cluster scores, and discrepancies between scores in the cognitive and ability areas. The WJ III NU computer scoring program offers options for interpreting intra-individual profiles of cognitive abilities and

achievement as well as ability-achievement discrepancies. The Woodcock Interpretation and Instructional Interventions Program software provides assistance with test interpretation by linking assessment results to evidence-based interventions and report writing.

Training Support: The publisher makes available training videos and workbooks and offers national and regional group training sessions as well as individual training sessions. Technical support is also available by telephone and online.

Adaptations/Special Instructions for Individuals with Disabilities: The Examiner's Manuals describe accommodations for testing individuals with various difficulties and impairments (including attention, behavioral, reading, hearing, visual, and physical disabilities).

Report Preparation Support: The Woodcock Interpretation and Instructional Interventions Program software provides assistance with test interpretation by linking assessment results to evidence-based interventions and report writing.

References:

Bradley-Johnson, Sharon, and Gokce Durmusoglu. "Evaluation of Floors and Item Gradients for Reading and Math Tests for Young Children." *Journal of Psychoeducational Assessment*, vol. 23, no. 3, 2005, pp. 262-278.

Edwards, Oliver, and Thomas Oakland. "Factorial Invariance of Woodcock-Johnson III Scores for African Americans and Caucasian

Americans." *Journal of Psychoeducational Assessment*, vol. 24, no. 4, 2006, pp. 358-366.

Krasa, Nancy. "Is the Woodcock-Johnson III a Test for All Seasons? Ceiling and Item Gradient Considerations in Its Use with Older Students." *Journal of Psychoeducational Assessment*, vol. 25, no. 1, 2007, pp. 3-16.

Mather, Nancy, and Richard W. Woodcock. *Examiner's Manual. Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing, 2001.

Mather, Nancy, and Richard W. Woodcock. *Examiner's Manual. Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing, 2001.

McGrew, Kevin S., Fredrick A. Schrank, and Richard W. Woodcock. *Technical Manual. Woodcock-Johnson III Normative Update*. Rolling Meadows, IL: Riverside Publishing, 2007.

McGrew, Kevin S., and Richard W. Woodcock. *Technical Manual. Woodcock-Johnson III*. Itasca, IL: Riverside Publishing, 2001.

Munoz-Sandoval, A.F., Richard W. Woodcock, Kevin S. McGrew, and Nancy Mather. *Bateria III Woodcock-Muñoz: Pruebas De Aprovechamiento*. Itasca, IL: Riverside Publishing, 2005a.

Munoz-Sandoval, Ana F., Richard W. Woodcock, Kevin S. McGrew, and Nancy Mather. *Bateria III Woodcock-Muñoz*. Itasca, IL: Riverside Publishing, 2005b.

Munoz-Sandoval, Ana F., Richard W. Woodcock, Kevin S. McGrew, and Nancy Mather. *Batería III Woodcock-Muñoz: Pruebas De Habilidades Cognitivas*. Itasca, IL: Riverside Publishing, 2005c.

Schrank, F.A., and Richard W. Woodcock. *WJ III Normative Update Compuscore and Profiles Program (Version 3.0) [Computer Software]*. *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside Publishing, 2007.

Schrank, Fredrick A., Kevin S. McGrew, Mary L. Ruef, Criselda G. Alvarado, Ana F. Munoz-Sandoval, and Richard W. Woodcock. "Overview and Technical Supplement (Batería III Woodcock-Muñoz Assessment Service Bulletin No. 1)." Itasca, IL: Riverside Publishing, 2005.

Schrank, Fredrick A., Barbara J. Wending, and Richard W. Woodcock. *Woodcock Interpretation and Instructional Interventions Program*. Rolling Meadows, IL: Riverside Publishing, 2007.

Woodcock, Richard W., K.S. McGrew, and Nancy Mather. *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside Publishing, 2007.

Woodcock, Richard W., Kevin S. McGrew, and Nancy Mather. *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing, 2007.

Woodcock, Richard W., Kevin S. McGrew, and Nancy Mather. *Woodcock-Johnson III Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside Publishing, 2007.

Woodcock, Richard W., Kevin S. McGrew, Nancy Mather, and Fredrick A. Schrank. *Woodcock-Johnson III Diagnostic Supplement to the Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside Publishing, 2007.

Woodcock, Richard W., Kevin S. McGrew, Fredrick A. Schrank, and Nancy Mather. *Woodcock-Johnson III Normative Update*. Rolling Meadows, IL: Riverside Publishing, 2007.

Woodcock, Richard W., Fredrick A. Schrank, Nancy Mather, and Kevin S. McGrew. *Woodcock-Johnson III Tests of Achievement, Form C/Brief Battery*. Rolling Meadows, IL: Riverside Publishing, 2007.

Parenting, the Home Environment, and Parent Well-Being Instruments



Parenting, the Home Environment, and Parent Well-Being Instruments^a

Instrument	Screening or Assessment	Domain					Age Range	Assessment Type	Initial Material Cost	Reliability	Validity	Norming Sample	Ease of Administration & Scoring
		H	P	PE	F	O							
Adult-Adolescent Parenting Inventory-2	S		X		X	X	13+ years	3	2	3	3	3	2
Beck Anxiety Inventory	S			X			17-80 years	3	1	3	3	2	2
Beck Depression Inventory-Second Edition	S			X			13-80 years	3	1	3	3	2	2
Center for Epidemiologic Studies Depression Scale	S			X			18+ years	3	1	2	3	1	2
Child Abuse Potential Inventory	S		X	X	X		Unspecified	3	2	3	2	1	2
Composite International Diagnostic Interview	A			X			Adults	3	2	3	3	1	3
Conflict Tactics Scales, Parent-Child Version	S		X				Unspecified	3	1	2	1	1	2
Confusion, Hubbub, and Order Scale	S	X					Unspecified	3	1	3	2	1	2
Early Head Start National Evaluation Questionnaires	NA		X										
Edinburgh Postnatal Depression Scale	S			X			16+ years	3	1	1	3	1	2
Family Environment Scale	A	X					Unspecified	3	3	3	1	2	2
Family Map of the Parenting Environment of Infants and Toddlers (IT-Family Map) and Family Map of the Parenting Environment in Early Childhood (EC-Family Map)	A	X	X	X	X		0-3 years, IT-Family Map and 3-5 years, EC-Family Map	2,3	1	1	1	1	2
Family Needs Scale	S				X		Unspecified	3	1	3	2	1	2
Family Resource Scale	S				X		Unspecified	3	1	3	3	1	2
Family Support Scale	S				X		Unspecified	3	1	2	2	1	2
The Home Observation for Measurement of the Environment Inventory for Infants/Toddlers (IT-HOME) and Early Childhood (EC-HOME)	A	X	X				0-3 years, IT-HOME and 3-6 years, EC-HOME	2,3	1	3	2,3	1	2
Infant-Toddler and Family Instrument	S	X	X			X	6-47 months	1,2	1	1	1	1	3
Kempe Family Stress Inventory	S		X				Unspecified	1	1	3	2	1	3
Knowledge of Infant Development Inventory	S		X				Unspecified	1	1	3	2	1	3
Nursing Child Assessment Teaching Scales	A		X				0-36 months	2	2	3	2	2	3

^a The information included in this table was drawn from the manuals or other resources available from the authors and publishers of the instruments. Individual users may have different experiences.

Parental Modernity Scale	A	X		Adults	3	1	3	3	1	2
Parenting Alliance Measure	A	X	X	Adults	3	2	3	3	2	2
Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO)	A	X		1-3 years	2	1	3	2	1	3
Parenting Stress Index	S	X		0-12 months	1	2	3	3	2	3
Support Functions Scale	S		X	Unspecified	3	1	2	2	1	2

KEY

Domains

H = Home environment
P = Parenting, parent-child relationship
PE = Parent mental health
F = Family functioning, support
O = Other

Assessment Type

1 = Direct assessment
2 = Observation
3 = Parent/ self report

Initial Material Cost

1 = Under \$100
2 = \$100 to \$200
3 = More than \$200

Reliability

1 = None described
2 = Under .65
3 = .65 or higher

Validity

1 = None described
2 = Under .5 for concurrent; under .4 for predictive
3 = .5 or higher for concurrent; .4 or higher for predictive

Validity

1 = None described
2 = Under .5 for concurrent; under .4 for predictive
3 = .5 or higher for concurrent; .4 or higher for predictive

Norming sample

1 = None described
2 = Older than 15 years, not nationally representative or representative of EHS population
3 = Normed within past 15 years, nationally representative or representative of EHS population

Ease of administration and scoring

1 = Not described
2 = Self-administered or administered and scored by someone with basic clerical skills
3 = Administered and scored by a highly trained individual

ADULT-ADOLESCENT PARENTING INVENTORY (AAPI-2), 1999

Authors:

Stephen J. Bavolek and Richard G. Keene

Publisher:

Family Development Resources, Inc.

800-688-5822

www.nurturingparenting.com/

-or-

www.familydev.com

Initial Material Cost:

The AAPI-2 complete kit (includes the handbook, test forms A and B, Scoring Stencil for forms A and B, profiles (pkg. of 100), worksheets (pkg. Of 100): \$122. Complete kit also available on CD-ROM for \$186.

Representativeness of Norming Sample:

A non-randomly selected sample (that the authors describe as nationally representative) of adolescents and adults (abusive and non-abusive adults, abused and non-abused adolescents, and teen parents) referred by agencies from around the country using the original AAPI participated in the standardization of the AAPI-2.

Languages:

English and Spanish

Type of Assessment:

Self-report

Age Range and Administration Interval:

Persons ages 13 and older

Personnel, Training, Administration, and Scoring Requirements:

An individual can learn to score the assessment by reading the manual. No training is necessary. Approximately 20 minutes to administer. Written at a 5th grade reading level.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher)

Validity: 3 (.5 or higher for concurrent/discriminant)

Norming Sample Characteristics: 3 (normed within the past 15 years, nationally representative)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Adult-Adolescent Parenting Inventory (AAPI-2) is a 40-item questionnaire used to assess the parenting attitudes and child rearing practices of adolescents and adults. The purpose of the inventory is to determine the degree to which respondents agree or disagree with parenting behaviors and attitudes known to contribute to child abuse and neglect. Responses are given on a five-point Likert scale ranging from Strongly Agree, Agree, Uncertain, Disagree, to Strongly Disagree. Responses provide a standard for risk in five parenting constructs known to contribute to the maltreatment of children: (1) inappropriate parental expectations, (2) inability

to demonstrate empathy towards children's needs, (3) strong belief in the use of corporal punishment, (4) reversing parent-child family roles, and (5) oppressing children's power and independence. The AAPI-2 comes in two alternate forms—A and B—to reduce the practice effect when repeating the inventory in a short time period.

Uses of Information: Responses to the AAPI-2 permit the identification of high-risk child rearing and parenting practices that could lead to physical or emotional abuse or neglect of children. In addition, the AAPI-2 is used to (1) provide pretest and posttest data to measure treatment

effectiveness, (2) assess the parenting and child rearing attitudes of parents and adolescents prior to parenthood, (3) design specific treatment and intervention parenting education programs, (4) design nurturing experiences for parents and adolescents whose attitudes indicate a high risk for child maltreatment, and (5) screen foster parent applicants, child care staff, and volunteers for education and training purposes.

Reliability: (1) Internal reliability: Reliability coefficients for the five parenting constructs using the Spearman-Brown formula ranged from .83 to .93 on Form A, .80 to .93 on Form B, and .87 to .96 on Forms A and B combined. The Cronbach alphas ranged from .80 to .92 on both Forms A and B and .86 to .96 on Forms A and B combined.

Validity: (1) Content validity: Statements made by parents about children formed the basis of the inventory items. Professionals in the helping fields assigned items to one of the five parenting constructs and assessed items' suitability for a Likert scale. (2) Construct validity: The authors provide factor analysis results that provide evidence for five underlying factors. (3) Criterion-related validity: A comparison between a group of abusive parents and a group of non-abusive parents (1,985 total sample size) found that abusive parents had mean scores on each of the parenting constructs that were statistically significantly lower than non-abusive parents. In general, males were also found to have lower scores than females, but there was no parenting-gender interaction effect. The authors provide evidence that the AAPI-2 discriminates between

abusive and non-abusive parents in samples of adults and in sample of adolescents.

Method of Scoring: Scoring is completed by placing a stencil over the test items and recording the numerical value of each response (1-5 points). The numerical values are recorded on the profile worksheet for each of the five subscales. The values are summed to obtain the subscale total raw score. The AAPI-2 worksheet and table of norms located in the Handbook convert total raw scores to standard scores for developing a risk profile on the worksheet. The respondent's attitudes in each of the five sub-scales can be compared with the parenting and child rearing attitudes of parents or adolescents. Tables convert raw scores to standard ten (sten) scores by gender (male or female), parental status (parents or non-parents), and age (adults or adolescents).

Interpretability: The standard scores are plotted on the AAPI Parenting Profile, which provides an index of risk for abusive and/or neglecting behaviors. The sten scores on the Profile sheet range from 1 to 10. Low sten scores (1 to 4) generally indicate a high risk for practicing known abusive parenting practices; mid-range scores (4 to 7) represent the parenting attitudes of the general population; and high sten scores (7 to 10) indicate the expressed parenting attitudes reflect a nurturing, non-abusive parenting philosophy. The manual provides instructions for interpreting the scale's total score and the subscale scores.

Training Support: Training workshops and training assistance is available. Call 828-681-8120 or send an email to fnc@nurturingparenting.com.

Adaptations/Special Instructions for Individuals with Disabilities: The assessment can be administered orally to non-readers.

Report Preparation Support: None described beyond the profiles.

References:

Bavolek, Stephen J. and Richard G. Keene. Adult-Adolescent Parenting Inventory AAPI-2 Administration and Development Handbook. Park City, UT: Family Development Resources, Inc., 1999.

BECK ANXIETY INVENTORY (BAI), 1993

Authors:

Aaron T. Beck and Robert A. Steer

Publisher:

The Psychological Corporation

(800) 228-0752

[http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8018-](http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8018-400&Mode=summary)

400&Mode=summary

Initial Materials:

Complete Kit: \$66 (Includes Manual and 25 Record Forms) Spanish Record Forms (25/package): \$35

Representativeness of Norming Sample:

Not nationally representative (the three normative samples of psychiatric outpatients were drawn from consecutive routine evaluations at the Center for Cognitive Therapy in Philadelphia, Pennsylvania. The total sample size was 1,086.)

Languages:

English and Spanish

Type of Assessment:

Self-report

Age Range and Administration Interval:

17 through 80 years

Personnel, Training, Administration, and Scoring

Requirements: The BAI may be administered and scored by paraprofessionals, but it should be used and interpreted only by professionals with appropriate clinical training and experience. The BAI requires 5 to 10 minutes to complete when it is self-administered and 10 minutes when it is orally administered.

Scoring takes 5 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (.65 or higher)

Validity: 3 (.5 or higher for concurrent)

Norming Sample Characteristics: 2 (not nationally representative, normed within past 15 years)

Ease of Administration and Scoring: 2 (self-administered; scored by someone with basic clerical skills)

Description: The Beck Anxiety Inventory (BAI) is a 21-item scale that measures the severity of self-reported anxiety in adults and adolescents. It consists of descriptive statements of anxiety symptoms which are rated on a 4-point scale with the following correspondence: "Not at all" (0 points); "Mildly; it did not bother me much" (1); "Moderately; it was very unpleasant, but I could stand it" (2); and "Severely; I could barely stand it" (3).

Uses of Information: The Beck Anxiety Inventory was specifically designed to reduce the overlap between depression and anxiety scales by

measuring anxiety symptoms shared minimally with those of depression.

Reliability: (1) Internal consistency (Cronbach's alpha) ranged from .92 to .94 for adults. The alphas for the Diagnostic and Statistical Manual of Mental Disorders, Third Edition—Revised (DSM-III-R) anxiety disorder groups ranged from .85 to .93. (2) Test-retest reliability (1-week interval): .75. Reliability of the BAI for adolescents has not been directly tested.

Validity: (1) Concurrent validity: the correlation with the Hamilton Anxiety Rating Scale—Revised was .51. The correlation with the

anxiety subscale of the Cognition Check List, which measures the frequency of dysfunctional cognitions related to anxiety, was also .51. The BAI is also significantly correlated with the Trait (.58) and State (.47) subscales of the State-Trait Anxiety Inventory (Form Y) and with the mean 7-day anxiety rating (.54) of the Weekly Record of Anxiety and Depression. Validity of the BAI for adolescents has not been directly tested.

Method of Scoring: The BAI total score is the sum of the ratings for the 21 symptoms. Each symptom is rated on a 4-point scale ranging from 0 to 3. The maximum score is 63 points. The Beck Computer Scoring (BCS) program also scores and interprets the BAI.

Interpretability: According to the 1993 Revisions of the BAI manual, total scores of 0 to 7 reflect “Minimal level of anxiety”; scores of 8 to 15 indicate “Mild anxiety”; scores of 16 to 25 reflect “Moderate anxiety”; and scores of 26 to 63

indicate “Severe anxiety.” Because a BAI total score yields only an estimate of the overall severity of anxiety being described by a person, the clinician interpreting the score should consider other aspects of the individual’s psychological functioning.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: If an examinee needs help to complete the inventory, instructions are given in the manual for how to administer the inventory orally.

Report Preparation Support: None described.

References:

Beck, Aaron T. and Robert A. Steer. Beck Anxiety Inventory Manual. San Antonio, TX: The Psychological Corporation Harcourt Brace & Company, 1993.

BECK DEPRESSION INVENTORY—SECOND EDITION (BDI-II), 1996

Authors:

Aaron T. Beck, Robert A. Steer, and Gregory K. Brown

Publisher:

The Psychological Corporation
(800) 228-0752
www.psychcorp.com

Initial Material Cost:

Complete kit: \$65 (includes manual and 25 record forms)
Beck InterpreTrak: \$50

Representativeness of Norming Sample:

Clinical and non-clinical sample of 500 outpatients from two urban and two rural psychiatric institutes

Languages:

English and Spanish (record forms only)

Type of Assessment:

Self-report

Age Range and Administration Interval:

13 through 80 years

Personnel, Training, Administration, and Scoring Requirements:

Requires only a few minutes to familiarize oneself with the inventory, 5 to 10 minutes to administer, and a few minutes to score.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: Internal consistency and test-retest

Reliability: 3 (.65 or higher)

Validity: 3 (Concurrent validity mostly above .5)

Norming Sample Characteristics: 2 (conducted within past 15 years, not nationally representative)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Beck Depression Inventory (second edition) is a self-administered tool for screening and assessing the severity of depression in adolescents and adults. Twenty-one items assess the intensity of depression in diagnosed patients as well as detect possible depression in normal population. Each item is a list of four statements arranged in increasing severity about a particular symptom of depression. This version of the BDI is in compliance with DSM-IV criteria for depression, and the age range covered has been expanded to 13 to 80 years of age.

Uses of Information: The BDI-II is an assessment of the severity of depression in psychiatrically diagnosed adults and adolescent patients aged 13 and older. It was developed as an indicator of the presence and degree of symptoms

correlated with depression as defined in the DSM-IV (including suicidality), not as an instrument for specifying a clinical diagnosis. It is also used extensively to monitor therapeutic progress.

Reliability: (1) Internal consistency (Cronbach's alpha) is .92 for clinical patients and .93 for non-clinical individuals. (2) Test-retest reliability is .93 (only for a small subsample of outpatients, tested with a one week lapse).

Validity: (1) Concurrent validity: two comparisons between BDI-II and its previous version resulted in correlations of .93 and .84, the latter using the take-home form. Other tests found BDI-II to be correlated with the Beck Hopelessness Scale (.68), Scale for Suicide Ideation (.37), Beck Anxiety Inventory (.60),

Hamilton Psychiatric Rating Scale for Depression-Revised (.71), and Hamilton Rating Scale for Anxiety - Revised (.47).

Method of Scoring: Most items on the BDI-II are rated on a 4-point scale ranging from 0 to 3. Several items have seven response options to discern differences in behavior or motivation. The BDI-II is scored by adding the ratings for the 21 items. The maximum total score is 63.

Interpretability: The interpretation of the scores should be done by professionals who have appropriate training and experience. Clinical interpretation of total scores uses the following guidelines: 0 to 13 (minimal depression),¹ 14 to 19 (mild depression), 20 to 28 (moderate depression), and 29 to 63 (severe depression). The Beck InterpreTrak software package offers a quick analysis of results for all of the Beck Scales (Depression, Anxiety, Hopelessness, and Suicidal Ideation) and then summarizes results in a single interpretive report with insights from Aaron T. Beck, M.D. IntepreTrak also helps monitor progress by generating longitudinal graphs and outcome ratings for each patient.

Training Support: Minimal training is required for administering or scoring the scale.

Adaptations/Special Instructions for Individuals with Disabilities: For individuals with reading or concentration difficulties, the items may be read aloud by the examiner. The manual includes instructions for both oral and self-administration. The manual also includes brief guidelines on how to help patients with severe depression understand the range of responses to the questions.

Report Preparation Support: The InterpreTrak software (available in CD-ROM or diskette) produces a comprehensive interpretive report.

References:

Beck, Aaron T., Gregory K. Brown, and Robert A. Steer. Beck Depression Inventory-II (BDI-II). San Antonio, TX: The Psychological Corporation, 1996.

Beck, Aaron T. Beck InterpreTrak. San Antonio, TX: The Psychological Corporation, 2000.

¹ Note that some research has shown that women who scored 0 or 1 tend to exhibit similar behaviors to high scoring women when observed in parent-child play. This has been attributed to denial—healthy people experience and endure at least some symptoms of depression.

CENTER FOR EPIDEMIOLOGIC STUDIES DEPRESSION SCALE (CES-D), 1985

Authors:

L.S. Radloff

Publisher:

National Institute of Mental Health

Initial Material Cost:

None

Representativeness of Norming Sample:

Scale not normed.

Languages:

English

Type of Assessment:

Self-report

Age Range and Administration Interval:

Intended for individuals over 18

Personnel, Training, Administration, and Scoring Requirements:

No training is required. The scale takes about 10 minutes to complete, and only a few minutes to score.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: Internal consistency: 3 (.65 or higher); Test-retest: 2 (< .65)

Validity: Concurrent: 3 (.5 or higher)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Center for Epidemiological Studies-Depression Scale (CES-D) is a 20-item instrument that can be self-administered or administered with minimal involvement by an interviewer. The instrument was developed by the National Institute of Mental Health to detect major or clinical depression in the general (nonpsychiatric) adult population (i.e., persons older than 18), specifically the frequency and duration of cognitive, affective, and behavioral depressive symptoms (within the past week).

Uses of Information: The CES-D is used for initial screening of symptoms related to depression or psychological distress. However, because the CES-D does not assess the full-range of depression symptoms (for example, it does not assess suicidality) and because it assesses the occurrence of the symptoms during the past week, users are cautioned against relying on the CES-D exclusively. It has also been used extensively for

research purposes to investigate levels of depression among the nonpsychiatric population.

Reliability: (1) Internal consistency reliability (Cronbach's alpha) ranged from .84 to .90 in field studies. (2) Test-retest reliability: Ranges from .51 to .67 in 2- to 8-week intervals and .41 to .54 in 3- to 12-month intervals.

Validity: (1) Concurrent validity: studies have examined the degree to which CES-D scores are in agreement with other measures of depression. These studies found CES-D to have correlations ranging from .50s to .80s with the Hamilton rating scale, .30s to .80s with the Raskin rating scale, .40s to .50s with the Lubin Depression Adjective Checklist, .60s and .20s, respectively, with the Bradburn Affect Balance Scale's Negative Affect and Positive Affect Scales, .50s with the Langner scale and .43 with the Cantril life satisfaction ladder. Discriminant validity tests found CES-D to be less successful in differentiating between

depression and other types of emotional responses, such as anger, fear, and boredom.

Method of Scoring: Respondents indicate the frequency or duration of time (in the past week) during which they have experienced certain feelings/situations. They circle a number between 0 and 3; 0 indicates that the situation occurred “rarely or none of the time” (less than 1 day), 1 indicates “some or a little of the time” (1 to 2 days), 2 indicates “occasionally or a moderate amount of time” (3 to 4 days), and 3 indicates “most or all of the time” (5 to 7 days). After adjusting the scores for the four positive-feature items, the item scores are summed to obtain the total scale score.

Interpretability: The possible range of total scores is from 0 to 60, with higher scores indicating greater distress. Radloff, the author of the scale, suggests that a total score of 16 be used as the cutoff to indicate “case” depression. However, other studies have suggested that scores of 0 to 15.5 be interpreted to indicate that an individual is “not depressed”, 16 to 20.5 to indicate “mild depression”, 21 to 30.5 to indicate “moderate depression”, and 31 or higher to

indicate “severe depression”. It is suggested that the scale be used only as an indicator of symptoms relating to depression, not as a means to clinically diagnose depression. Therefore, higher scores on the CES-D scale may indicate a need for further clinical tests/screenings. However, because of the CES-D’s limitations, a low score does not necessarily indicate the absence of clinical depression.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Devins, Gerald M. and Carolee M. Orme. “Center for Epidemiologic Studies Depression Scale.” In *Test Critiques*, edited by D.J. Keyser and R.C. Sweetland. Kansas City, MO: Test Corporation of America, 1985.

Radloff, L.S. “The CES-D Scale: A Self-Report Depression Scale for Research in the General Population.” *Applied Psychological Measurement*, vol. 1, 1977, pp. 385-401.

THE CHILD ABUSE POTENTIAL INVENTORY (CAP), SECOND EDITION, 1986

Authors:

Joel S. Milner

Publisher:

Psytec Corporation (815) 758-1415

Initial Material Cost:

CAP Inventory Manual: \$30

Interpretive Manual: \$20

Package of ten tests: \$16 (also available in packages of 25, 50, and 100)

Hand-scoring templates: \$50

CAPSCORE computer scoring program: \$195
(software is currently being updated)

Representativeness of Norming Sample:

Not nationally representative

Languages:

English and Spanish

Type of Assessment:

Parent or caregiver self-report

Age Range and Administration Interval:

Not applicable

Personnel, Training, Administration, and Scoring Requirements:

A trained nonprofessional under the supervision of a qualified professional can administer the Inventory. However, interpretation should be done by a professionally trained social worker, counselor, psychologist, or other professional with advanced training in assessment and test interpretation. This inventory requires a 3rd grade reading level. It takes 20 minutes to administer. Scoring can be done by computer or by hand.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (.65 or higher) for 77-item CAP abuse scale

Validity: 2 (.5 or higher for concurrent and < .4 for predictive)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered, scored by a highly trained individual or computer program)

Description: The CAP Inventory is a 160-item questionnaire designed to assist in screening male and female parents or primary caregivers who are suspected of physical child abuse. The Inventory (Form VI) contains a total of 10 scales. The primary clinical scale is the 77-item physical child abuse scale. This abuse scale can be divided into six factor scales: distress, rigidity, unhappiness, problems with child and self, problems with family, and problems from others. In addition, the CAP Inventory contains three validity scales: the lie scale, the random response scale, and the inconsistency scale. The validity scales are used in various combinations to

produce three response distortion indexes: the faking-good index, faking-bad index, and random response index. This instrument should always be used in conjunction with evaluation data from other sources, including interviews and other test data.

Uses of Information: The CAP Inventory is intended to assist in the screening of suspected physical child abuse cases in social services agencies and similar settings. It can also be used as a screening tool for the selection of individuals who are at increased risk for physical child abuse,

to assess clients prior to treatment, or for treatment or program evaluation purposes.

Reliability: (1) Split-half reliability: Split-half reliabilities ranging from .93 to .98 and Kuder Richardson-20 coefficients ranging from .85 to .96 have been reported for different gender, age, educational level, and ethnic groups.

Validity: (1) Construct validity: The CAP abuse score is positively correlated (.48) with the amount of physical abuse in childhood. (2) Predictive validity: a significant correlation of .34 was found between abuse scores and subsequent confirmed reports of abuse and neglect.

Method of Scoring: Each item is answered in a forced-choice, agree-disagree format. Scoring can be done by hand or by computer using a computer-scoring program (CAPSCORE). The hand scoring approach uses a series of transparent scoring templates to generate the scale scores. The name of each CAP Inventory scale and the associated items to be scored are indicated on each template. For the Abuse Scale and six factor scales, weighted scores are then summed and scale scores are produced. For the Validity Scales, a nonweighted scoring procedure is used.

The Response Distortion Index Scores are determined using the raw score totals of different pairs of the individual validity scales. A non-weighted scoring procedure is used for Special

Scale Scores. To avoid errors, it is recommended that the CAPSCORE program be purchased and used to score the Inventory. It computes all of the above scores automatically.

Interpretability: Interpretation rules for the validity scales, the response distortion indexes, the CAP abuse scale, and the six abuse factor scales are outlined in the Technical and the Interpretive Manuals. Cut-off scores are listed in the manual. The author recommends using a cut-off score of 116, or 215 out of the possible 0 to 400 points, depending on the sensitivity and specificity required. The manual specifies that the 77-item abuse scale score, not individual factor scores, should be employed for the screening of physical child abusers.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Milner, Joel S. *The Child Abuse Potential Inventory: Manual (Second Edition)*. DeKalb, IL: Psytec, 1986.

Milner, Joel S. *An Interpretive Manual for The Child Abuse Potential Inventory*. Webster, NC: Psytec, 1990.

COMPOSITE INTERNATIONAL DIAGNOSTIC INTERVIEW (CIDI), 1997

Authors:

World Health Organization

Publisher:

World Health Organization
www.who.int/msa/cidi/

Initial Material Cost:

Complete Package (includes lifetime and 12-month interviews, plus interviewer's and trainer's manual, Probe Flow Chart, and question specifications. Also includes the scoring and data entry programs for both interviews.): \$150;

CIDI Auto (The computerized version of the CIDI): \$500

Representativeness of Norming Sample:

None described

Languages:

English, Spanish, French, Italian, Dutch, German, and Portuguese¹

Type of Assessment:

Self report

Age Range and Administration Interval:

Adults who can read and write

Personnel, Training, Administration, and Scoring Requirements:

Administration time is 75 minutes for the regular form; 20 minutes for the short form. The instruments can be administered by a trained nonclinician and scored by a clerical individual. Computer scoring takes 20 minutes.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: Inter-rater: 3 (kappa .65 and higher), Test-retest: 3 (kappa mostly .65 and higher)

Validity: Concurrent: 3 (mostly .5 and higher)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

¹ Other language versions are in preparation.

Description: The Composite International Diagnostic Interview (CIDI) is a comprehensive, fully-structured psychiatric diagnostic interview designed to be used by trained nonclinician interviewers to diagnose more than 40 mental disorders among adults from different cultures according to the definitions and criteria of both the International Classification of Diseases, 10th edition) (ICD-10) and the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV) diagnostic systems for lifetime, last year, last 6 months, last month, and last 2 weeks. The CIDI is available in lifetime and 12-month versions, and in both paper and pencil and computer-administered forms. The latter version

is suitable for self-administration by cooperative subjects. During a CIDI interview, respondents are asked closed-ended questions about symptoms of psychiatric disorders. Positive responses to some of the symptom questions are followed by questions from the Probe Flow Chart that determine whether the symptom is a possible psychiatric symptom (that is, it is clinically significant and is not due to medication, drugs or alcohol or to a physical illness or injury). Negative responses to symptom questions will often lead to later questions being skipped. If enough symptoms have been endorsed, and these symptoms occur in a pattern that suggests a diagnosis might be present, respondents are asked

about the onset and the recency of the particular cluster of symptoms that they have endorsed. In addition to the CIDI, a shortened form exists, CIDI-SF, used for the period of the past 12 months. Six DSM-IV mental disorders and two DSM-III-R substance disorders are evaluated with the CIDI-SF: major depression, generalized anxiety, specific phobia, social phobia, agoraphobia, panic attack, alcohol dependence, and drug dependence.

Uses of Information: The CIDI is used to determine whether or not a person is likely to suffer from a mental disorder.

Reliability: (1) Inter-rater reliability: An intra-class kappa of 1.00. (2) Test-retest: Test-retest (with a one-month interval) kappa coefficients for substance abuse disorders over the respondents' lifetime using the Munich CIDI ranged from .55 (drug abuse) to .83 (alcohol abuse). The Brazilian CIDI yielded test-retest (no test interval provided) kappa coefficients ranging from .61 to 1.00 on all psychiatric and substance abuse/dependency disorders, except for alcohol abuse, which had a coefficient of .35. The kappas for simple phobia, social phobia, and agoraphobia over the respondents' lifetimes were .46, .47, and .63, respectively, and, for generalized anxiety disorder, it was .53.

Validity²: Concurrent validity (referred to as "concordance validity" by authors): A comparison between the CIDI and the Structured Clinical

² Many of these studies were conducted using an earlier version of the CIDI.

Interview for DMS-III-R (SCID) on simple phobia, social phobia, and agoraphobia disorders yielded kappa coefficients of .45, .62, and .63, respectively, and .35 on the lifetime generalized anxiety disorder. A comparison between the CIDI and the clinical DSM-III-R criteria checklist produced kappa coefficients of .84, .83, and .76 for depressive, psychoactive substance, and anxiety disorders, respectively, and .78 for all disorders. The canonical correlation coefficients between the CIDI and the Schedules for Clinical Assessment in Neuropsychiatry on anxiety and depressive disorders were .66 for lifetime disorders and .69 for current disorders. Two studies, one that compared the CIDI-Auto with those of psychiatrists and the other that compared CIDI-Auto with pairs of clinicians, found that CIDI-Auto tended to identify more disorders than the mental professionals. The CIDI-Auto generated an average of 2.3 diagnoses of general disorders per subject compared to 1.3 diagnoses for psychiatrist and twice as many anxiety diagnoses than the clinicians. A kappa coefficient of .23 was obtained between the diagnoses of the CIDI and the psychiatrists. The CIDI-Auto sensitivity was above .85 for all anxiety disorders, except generalized anxiety disorder, which had a sensitivity of .29.³ Its specificity ranged from .47 to .99.⁴ The level of agreement between the CIDI-Auto and the clinicians, as measured by intraclass

³ Sensitivity is a measure of the instrument's ability to correctly identify persons with the disorder as having the disorder.

⁴ Specificity is a measure of the instrument's ability to identify persons who do not have the disorder as not having the disorder.

kappa, ranged from .02 to .81, with an overall kappa of .40.

Method of Scoring: Scoring can be done manually or by using the computerized version of the CIDI, known as CIDI-Auto, which is an SPSS-based program. The CIDI-SF is scored manually by summing the number of positive responses to symptoms the respondent reported. The manual for scoring the CIDI-SF is available at the CIDI Website.

Interpretability: No instructions were available for interpreting the CIDI. The CIDI-SF uses a probability-of-caseness score to indicate the likelihood that the respondent would meet the full diagnostic criteria if given the complete CIDI. Tables are used to convert CIDI-SF raw scores for each disorder into probability-of-caseness values that range from 0.0 to 1.0. Alternatively, the examiner can elect not to use the probability values and, instead, consider all probabilities greater than .50 to indicate that the respondent would be a CIDI case for that disorder.

Training Support: Administration of the interview requires training on skip patterns, on the use of the Probe Flow Chart, in assembling lists of the endorsed symptoms for the onset and

recency questions, and in the use of the data entry and scoring program. Training in administering this structured interview is conducted at nine WHO endorsed centers around the world. The WHO-CIDI Website URL is www.who.int/msa/cidi/. The cost is \$1000.

Adaptations/Special Instructions for Individuals with Disabilities: None

Report Preparation Support: None

References:

Robins, Lee N., John Wing, Hans Ulrich Wittchen, John E. Helzer, Thomas F. Babor, Jay Burke, Anne Farmer, Assen Jablenski, Roy Pickens, Darrel A. Regier, Norman Sartorius, Leland H. Towle. "The Composite International Diagnostic Interview: An Epidemiologic Instrument Suitable for Use in Conjunction With Different Diagnostic Systems and in Different Cultures." *Arch Gen Psychiatry*, Vol. 45, Dec. 1988, pp. 1069-1077.

World Health Organization. *Composite International Diagnostic Interview (CIDI), Core Version 2.1, Interviewer's Manual*. World Health Organization, January 1997.

CONFLICT TACTICS SCALES, PARENT-CHILD VERSION (CTSPC), 1998

Authors:

Murray A. Straus, Sherry L. Hamby, David Finkelhor, David W. Moore, & Desmond Runyan

Publisher:

Family Research Laboratory University of New Hampshire
(603) 862-1888
http://pubpages.unh.edu/~mas2/CTS_Application.htm

Initial Material Cost:

Conflict Tactics Scale (CTS) Handbook: \$25 (This 350-page manual includes all versions of the CTS, most of the currently available CTS series publications, Spanish translations of the original CTS—Forms N and R, and other papers. The article in Child Abuse And Neglect (paper CTS17 on the above website) is the only manual for the CTSPC. However, there are other relevant articles such as CTS24 and CTS28.)

Representativeness of Norming Sample:

Not normed.

Languages:

English and French translations exist for the CTSPC. Other forms of the CTS may have other translations.

Type of Assessment:

Parent report (self-administration), but can also be administered in person or over the telephone (direct parent assessment).

Age Range and Administration Interval:

Parents of children

Personnel, Training, Administration, and Scoring Requirements:

An individual with roughly a 6th grade reading level can complete the scales. The only training course is a 4-hour workshop that Dr. Straus runs every year in connection with an annual conference on family violence research. Administration time is 10 to 15 minutes if the entire scale is administered and 6 to 8 minutes if the supplemental questions are omitted.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 2 (under .65)

Validity: 1 (statistics not provided)

Norming Sample Characteristics: 1 (not described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Conflict Tactics Scales, Parent-Child Version (CTSPC) is intended to measure psychological and physical maltreatment and neglect of children by parents, as well as non-violent modes of discipline. It measures the extent to which a parent has carried out specific acts of physical and psychological aggression, regardless of whether the child was injured. Variables are measured on three scales: Non-Violent Discipline, Psychological Aggression, and Physical Assault, as well as supplemental scales that measure Weekly Discipline, Neglect, and Sexual Abuse.

Uses of Information: The CTSPC may be used as a screening tool for child maltreatment or for evaluating prevention and treatment of physical and psychological maltreatment of children.

Reliability: ¹(1) Internal reliability (Cronbach's alphas): Overall Physical Assault Scale: .55; Psychological Aggression: .60; Nonviolent Discipline: .70; Neglect Scale: .22;

¹ Previous versions of the CTS were tested for reliability and validity.

Severe Physical Assault Subscale: -.02. The authors attribute the low neglect and severe assault alphas to the infrequency of the events that make up the scales, thereby reducing the likelihood for high inter-item correlations. (2) Test-retest reliability is not yet available for the CTSPC. However, the test-retest reliability coefficients on the original CTS (test interval not specified) ranged from .49 to .80.

Validity:² The authors tested for construct validity by examining the direction of the relationship between subscale scores and demographic characteristics associated with child maltreatment, such as age of parent, age of child, race/ethnicity, and gender of parent. The directions of the relationships were consistent with previous findings.

Method of Scoring: Most of the scales can be scored four ways: (1) Annual prevalence, which measures whether one or more acts in the scale occurred during past year; (2) annual chronicity, which measures the number of times an act in a scale occurred among those who used that act; (3) ever prevalence, which measures if an act ever occurred; and (4) annual frequency, which measures the number of times an act occurred. To obtain the frequency, the midpoints for the response categories chosen by the participant are summed.

Interpretability: Normative tables for the CTSPC have not yet been developed.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Straus, M.A., S. L. Hamby, D. Finkelhor, D.W. Moore, and D. Runyan. "Identification of Child Maltreatment With the Parent-Child Conflict Tactics Scales: Development and Psychometric Data for a National Sample of American Parents." *Child Abuse & Neglect*, vol. 22, no. 4. 1998.

Straus, Murray A. "Scoring and Norms for the CTS2 and CTSPC." <http://pubpages.unh.edu/~mas2/CTS28.pdf>, 1998.

Straus, Murray A. "Child-Report, Adult-Recall, and Sibling Versions of the Revised Conflict Tactics Scales." <http://pubpages.unh.edu/~mas2/CTS24.pdf>, 1999.

² Previous versions of the CTS were tested for reliability and validity.

CONFUSION, HUBBUB, AND ORDER SCALE (CHAOS), 1995

Authors:

Adam P. Matheny, Jr., Theodore D. Wachs, Jennifer L. Ludwig, and Kay Phillips

Publisher:

Child Development Unit
Department Pediatrics
University of Louisville Health Service Center

Initial Material Cost:

None. The scale is available in the Journal of Applied Developmental Psychology article cited below.

Representativeness of Norming Sample:

Instrument is not normed.

Languages:

English

Type of Assessment:

Parent report

Age Range and Administration Interval:

Age of children not specified, but the assessment appears targeted for homes with infants and toddlers.

Personnel, Training, Administration, and Scoring Requirements:

A caregiver or parent who is literate can answer the 15 items. Scoring requires summing the responses given by the parent and takes under 5 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (.65 or higher)

Validity: 2 (concurrent under .5)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The CHAOS scale is a questionnaire filled out by parents that is designed to assess the level of confusion and disorganization in the child's home environment. The questionnaire consists of 15 statements, to each of which a parent or caregiver assigns a number between 1 and 4 that correspond to the following: 1 = Very much like your own home; 2 = Somewhat like your own home; 3 = A little bit like your own home; 4 = Not at all like your own home.

Uses of Information: The CHAOS scale screens for a chaotic home environment. High levels of chaos for at-risk children may warrant a more detailed environmental assessment to determine how and to what degree ongoing chaos is either compounding the effects of existing

biosocial risks or attenuating the impact of corrective intervention.

Reliability:¹ (1) Internal consistency (Cronbach's alpha): For the entire scale, .79. (2) Test-retest reliability (12-month interval): for the total test score, .74.

Validity:² (1) Concurrent validity: the CHAOS scale was compared with the physical and social environment codes in the Purdue Home Simulation Inventory (PHSI), which are completed by trained observers. The authors report that the correlations between the CHAOS scale and several of the PHSI social environment

¹ These results are based on an earlier version of the CHAOS Scale that used a true-false scoring system.

² These results are based on an earlier version of the CHAOS Scale that used a true-false scoring system.

codes were significant (physical interference (correlation =-.36), number of known objects named (correlation =-.38), and ignores bids (correlation =.45)), and together, the PHSI social environment codes explained 59 percent of the variance in the CHAOS scores. The correlations between the CHAOS scale and several of the PHSI physical environment codes also were significant (number of siblings (correlation =.55) and number of rooms per person (correlation =-.33)), and together the PHSI physical environment codes explained 39 percent of the variance in the CHAOS scores.

Method of Scoring: The statements are scored using a 4-point scoring system. A single score is derived from the CHAOS questionnaire by summing the responses for the 15 items. A

higher score represents characteristics of a more chaotic, disorganized, and hurried home.

Interpretability: The higher the score, the more chaotic a home is considered to be.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Matheny, Adam P., Jr., Theodore D. Wachs, Jennifer L. Ludwig, and Kay Phillips. "Bringing Order Out of Chaos: Psychometric Characteristics of the Confusion, Hubbub, and Order Scale." *Journal of Applied Developmental Psychology*, vol. 16, 1995, pp. 429-444.

EDINBURGH POSTNATAL DEPRESSION SCALE, 1987

Authors:

J.L. Cox, J.M. Holden, and R. Sagovsky

Publisher:

None.

Initial Material Cost:

None. A copy of the scale is available in Cox, J.L., Holden, J.M., & Sagovsky, R. "Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale." *British Journal of Psychiatry*, Vol. 150, 782-786, 1987.

Representativeness of Norming/Research Sample:

No norming sample

Languages:

English

Type of Assessment:

Self-report

Age Range and Administration Interval:

Women of childbearing age

Personnel, Training, Administration, and Scoring Requirements:

Administration time is 10 minutes; Scoring can be done in about 5 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described)

Validity: 3 (.5 or higher)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Edinburgh Postnatal Depression Scale (EPDS) is a measurement tool that is used to screen for depression during the postpartum (postnatal) period. Mothers underline the response items that most closely reflect her feelings during the past week.

Uses of Information: The EPDS is designed to detect women suffering from postnatal depression. It does not provide information on the severity of the depression. A respondent whose score is indicative of probable postnatal depression should have a comprehensive assessment.

Reliability: None described.

Validity: (1) Concurrent Validity: a validation study on British mothers found that a 12.5 cutoff score identified over 80 percent of the mothers with major depression and about 50 percent of the mothers with minor depression,

and had a sensitivity value of 67.7 percent. Another study found a score of 9.5 or higher to be more appropriate for identifying depression among Chinese mothers.

Method of Scoring: Responses are scored from 0 to 3 according to increased severity of the symptoms. Individual items are totaled to give an overall score.

Interpretability: A score of 12 or more on EPDS or an affirmative answer on question 10 (presence of suicidal thoughts) requires more thorough evaluation.

Training Support: None described, but none seems to be needed.

Adaptations/Special Instructions for Individuals with Disabilities: None described. However, the instrument can quite easily be administered in an interview format, if mental

or physical disabilities make it difficult for a respondent to complete the instrument.

Report Preparation Support: None described.

References:

Lee, DTS, and TKH Chung. "What should be done about postnatal depression in Hong Kong?" *Hong Kong Medical Journal*, Vol. 5, No. 1, pp. 39-42, March 1999.

Murray, Lynne and Andrew D. Carothers. "The Validation of the Edinburgh Post-natal Depression Scale on a Community Sample." *British Journal of Psychiatry* Vol. 157, pp. 288-290, 1990.

Warner, R., Appleby, L., Whitton, A., & Faraghen, B. "Demographic and obstetric risk factors for postnatal psychiatric morbidity." *British Journal of Psychiatry*, Vol. 168, 607-611, 1996.

FAMILY ENVIRONMENT SCALE (FES), 1994

Authors:

Rudolf H. Moos and Bernice S. Moos

Publisher:

Mind Garden
(650) 261-3500
www.mindgarden.com

Initial Material Cost:

Manual: \$56
Interpretative Report Forms: \$1 each or \$34 for 25 FES Item Booklets,
Expectations and Ideal Forms: \$2 each or \$48 for 25;
Real Form: \$1 each or \$32 for 25
Scoring Key: \$15 each
Self-Scorable Answer Sheets: \$1 each or \$41 for 25
Non-Paid Answer Sheets: \$1 each or \$16 for 25
Self-Scorable Preview Kit: \$57

Representativeness of Norming Sample:

Form R was normed on a sample of 1,432 normal families and 788 distressed families. The normal families were diverse in terms of geography, family type, race, and age. They also included 601 families that served as normal comparison groups in studies of alcoholic and depressed families.¹ The distressed families had members who were alcohol abusers, depressed or psychiatric patients, family clinic patients, on probation or parole, and adolescents or younger children in crisis situations. Form I was normed on a sample of 591 individuals from varied family types,

¹ See description section below for information on the different types of forms.

including normal and distressed individuals. No separate norming sample was drawn for Form E. Form E scores are normed using the Form R sample. Cross-cultural normative samples are available from the translated and culturally adapted versions of the FES.

Languages:

English, Arabic, Chinese, Dutch, Estonian, French, German, Hebrew, Hindi, Italian, Japanese, Korean, Marathi, Portuguese, Russian, Spanish, and Swedish.

Type of Assessment:

Child (11 years or older) and parent report on family environment.

Age Range and Administration Interval:

Not applicable. Focus is on the family environment.

Personnel, Training, Administration, and Scoring Requirements:

Instructions for administering the FES are self-explanatory and no training is required. It usually takes individuals 15 to 20 minutes to complete each of the three forms. It takes approximately 10 minutes to obtain raw scores for all 90 items.

Summary:

Initial Material Cost: 3 (> \$200)
Reliability: 3 (.65 or higher).
Validity: 1 (concurrent not available).
Norming Sample Characteristics: 2 (normed within past 15 years; diverse but not representative).
Ease of Administration and Scoring: 2 (self-administered; scored by someone with basic clerical skills)

Description: The Family Environment Scale (FES) is one of 10 Social Climate Scales, each assessing the climate in a different setting with 10 subscales organized into three dimensions—relationship, personal growth, and system maintenance. The FES measures family social environment using three forms with 90 true-false items: (1) the Expectations Form (Form E) for

information on expectations from a new family environment, (2) the Real Form (Form R) for information on perceptions of the current family environment, and (3) the Ideal form (Form I) for information on the preferred family environment. In addition, there is a 30-item pictorial children's version for use with children between the ages of 5 and 11. The FES is administered to family

members as a paper- and pencil-inventory with true or false answers.

Uses of Information: The FES can be used for the following purposes: (1) understanding problems in family functioning, (2) serving as a benchmark to evaluate the impact of an intervention, (3) providing feedback to families as a means to promote change, (4) evaluating how a family has been affected by a transition, life crisis or change (provided comparable information is available about the family prior to the event), (5) appraising and improving the family climate parents create, (6) strengthening families as cohesive units, (7) identifying risks for various problems, such as, depression, substance abuse, or family violence.

Reliability: (1) Internal consistency (Cronbach's alpha): the subscale alphas for Form R ranged from .61 for independence to .78 for cohesion, intellectual-cultural orientation, and moral-religious emphasis. No alphas were reported for Form I; however, the authors reported that they were similar to Form R alphas. (2) Test-retest reliability: The Form R subscale reliability coefficients ranged from .68 for independence to .86 for cohesion with a 2-month testing interval and .54 for independence to .91 for moral-religious with a 4-month testing interval.

Validity: The authors reported, as evidence of construct validity, studies that found results on the FES subscales to be consistent with the results on other instruments measuring the same construct and the lack of a relationship with results on instruments measuring different

constructs. These instruments included the Social Support Appraisals (SS-A; Vaux et al., 1986), the Social Support Questionnaire (Sarason, et al., 1987), the Locke-Wallace Marital Adjustment Scale (Waring et al., 1981), the Spanier Dyadic Adjustment Scale (DAS; Abbott & Brody, 1985), the Parental Bonding Instrument (Sarason, et al., 1987), the Family Assessment Device (FAD) and the Family Adaptability and Cohesion Evaluation Scales (FACES-II; Dickerson and Coyne, 1987), FACES-III (Edman, Cole, and Howard, 1990), the Structural Family Interaction Scale – revised (Perosa and Perosa, 1990), and the Family System Test (FAST; Feldman and Gehring, 1988), the Family Sculpture Test, and an adapted version of the Bowerman and Bahr Identification Scale (Russell, 1980). The authors did not report any statistics on the magnitude of the relationships.

Method of Scoring: Individuals complete subscale questions with true or false answers on separate answer sheets by placing an X in the appropriate column. The examiner then uses a template to score the responses by summing the number of X's in each column. When the FES is administered to more than one family member, the subscale raw scores for each family member are averaged in order to obtain the family's mean raw score for each subscale. Tables are provided to assist the clinician in converting the raw score to a standard score.

Interpretability: Subscale responses are compared to those of a group of normal families, using standard scores, which have a mean of 50 and standard deviation of 10. The manual

provides tables for converting Form R subscale and family incongruence raw scores into standard scores. The same table can be used to convert Form E raw scores into standard scores. Program staff with a basic knowledge of statistics can calculate equivalent percentiles based on the family's standard scores, the mean, and the standard deviation. The manual provides case studies to help interpret the results.

Training Support: None described

Adaptations/Special Instructions for Individuals with Disabilities: It is recommended that the FES be administered using tape-recorded or computerized instructions for those individuals who have short attention spans or cannot read at a sixth-grade level. It may also be helpful to administer the FES in individual interviews for poor functioning residents of treatment or residential care facilities. Some people will not be able to understand the questions, including children under the age of 11 and individuals who are mentally retarded, seriously impaired psychiatrically, or who suffer from a chronic brain disorder or cognitive dysfunction.

Report Preparation Support: A sample narrative report is provided in order to help clinicians interpret the results from the FES.

References:

Abbott, D.A. & G.H. Brody. The relation of child age, gender and number of children to the marital adjustment of wives. *Journal of Marriage and the Family*, vol. 47, 1985, pp. 77-84.

Dickerson, V.C. and J.C. Coyne. Family cohesion and control: A multitrait/multimethod study, *Journal of Marital and Family Therapy*, vol. 13, 1987, pp. 275-285.

Feldman, S.S. and T.M. Gehring. Changing perceptions of family cohesion and power across adolescence, *Child Development*, vol. 59, 1988, pp. 1034-1045.

Moos, Rudolf and Bernice Moos. *Family Environment Scale Manual: Development, Applications, Research, Third Edition*. Palo Alto: Consulting Psychologists Press, Inc., 1994.

Perosa, L.M. and S.L. Perosa. Convergent and discriminant validity for family self-report measures, *Educational and Psychological Measurement*, vol. 50, 1990, pp. 855-868.

Russell, C. A methodological study of family cohesion and adaptability, *Journal of Marital and Family Therapy*, vol. 6, 1980, pp. 459-470.

Sarason, B.R., et al. Interrelations of social support measures : Theoretical and practical implications. *Journal of Personality and Social Psychology*, vol. 52, 1987, pp. 813-832.

Vaux, A., et al. The Social Support Appraisals (SS-A) Scale : Studies of reliability and validity. *American Journal of Community Psychology*, vol. 14, 1986, pp. 195-219.

Waring, E.M., et al. Dimensions of intimacy in marriage. *Psychiatry*, vol. 44, 1981, pp. 169-175.

FAMILY MAP OF THE PARENTING ENVIRONMENT OF INFANTS AND TODDLERS (IT-FAMILY MAP) AND FAMILY MAP OF THE PARENTING ENVIRONMENT IN EARLY CHILDHOOD (EC-FAMILY MAP) (2010) (2011 Update)

Authors:

Leanne Whiteside-Mansell, Robert H. Bradley, Nicola Burrow, and Patti Bokony

Publisher:

University of Arkansas for Medical Sciences,
Department of Family & Preventive Medicine
(501) 686-7633
<http://familymedicine.uams.edu/FamilyMap>

Initial Material Cost: The IT- and EC-Family Map measures (Manual, Response Cards, and Observation Summary card) and manuals may be requested from developers (printing costs may apply).

Representativeness of Norming Sample:

No norming sample

Languages:

English, Spanish

Type of Assessment:

Parent interview and home observation

Age Range and Administration Interval:

Birth to 5 years. Recommended for administration twice a year during fall and spring home visits.

Personnel, Training, Administration, and Scoring

Requirements: Administered by Early Head Start and other early child care program staff members.

Assessors should spend 5 to 10 hours studying the training manual and practicing administration. Developers recommend attendance at a one-day training session, although 4-hour trainings are also available for specialists with master's degrees in areas such as social work and early childhood development. A training manual and video demonstration of the Family Map administration are available from the developer upon request. Web-based training materials are under development.

The Family Map takes about 1 hour to administer and score.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described for the IT-Family Map) and 3 (mostly over 80 percent agreement for EC-Family Map inter-rater reliability)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 2 (self-administered and scored by someone with basic training)

Description: The Family Map of the Parenting Environment of Infants and Toddlers (IT-Family Map) the Parenting Environment in Early Childhood (EC-Family Map) measure family and home risk factors, needs, and strengths for children birth to 3 and 3 to 5 years of age, respectively. The Family Map is a parent interview and home observation used during home visits with families in Early Head Start and non-Head Start early child care settings. Response cards are

available for assessors to prompt parents with possible responses for each question. Two related measures are under development: the Family Map of the Prenatal Environment for Mothers in Early Head Start programs and the Family Map for the Hearing-Impaired Child.

The Family Map aligns with Head Start objectives and the targeted components of family life instrumental to child well-being. Broadly, it

assesses physical and social conditions that children directly experience, family climate and context, and parental characteristics. The Family Map includes 11 subscales and an observation, with 114 (IT-Family Map) or 103 items (EC-Family Map). The subscales include Demographics, Routines, School Readiness, Monitoring, Environmental Safety, Family Cohesion (called Family Support and Conflict in the EC-Family Map), Discipline, Physical and Mental Health (called Healthy in the EC-Family Map), Basic Needs, Home and Car Safety, and Social Support and Integration (called Social Integration in the EC-Family Map). The observation at the end of the interview assesses Parenting Warmth and Discipline.

Other Languages: The developer has translated the forms and response cards for the EC-Family MAP into Spanish. Information on the norming sample, validity, reliability, and English language equivalence is unavailable.

Uses of Information: During routine home visits, Head Start professionals use the Family Map to identify strengths and concerns in family life related to children's adaptive functioning to identify and target family goals (Whiteside-Mansell et al. 2007) and enhance the provider-parent relationship (Whiteside-Mansell et al. 2008a). It may also be used in the Head Start self-assessment and monitoring process, the detection of program impacts based on comparisons between fall and spring home visit results (Whiteside-Mansell et al. 2010; Whiteside-Mansell et al. 2008b), and the identification of

areas to target for agency-wide intervention to reduce risk factors (e.g., food insecurity, physical safety issues, family conflict, harsh parenting practices, parental depression) and enhance protective factors associated with healthy development (e.g., availability of learning materials in the home, good monitoring and supervision, home safety). The IT-Family Map detects drug and substance abuse among household members and offers program members opportunities for support and resources; such resources may be particularly helpful to program providers because drug and alcohol use may be difficult to discuss and therefore identify (Bokony 2010).

Reliability:

1. Internal consistency reliability: no information available.
2. Test-retest reliability: no information available.
3. Inter-rater reliability: teachers and research assistants implementing the IT-Family Map are currently being evaluated for inter-rater reliability (personal communication with Whiteside-Mansell 2011). For the EC-Family Map, teachers interviewed 20 parents, and data collectors interviewed the same parents two weeks later and achieved over 80 percent agreement among most of the 59 indicators (i.e., groups of items) assessed (Whiteside-Mansell et al. 2007). Three subscales contained indicators with less than 80 percent agreement. Within the Family Cohesion and

Routines subscale, research assistants identified twice the number of families at risk for parenting stress and routines related to daily activities. Developers are enhancing Family Map training materials to address the low inter-rater reliability of the parenting stress indicators. Within Basic Needs, research assistants identified two additional at-risk families for food quality and poison accessibility. Within Environmental Safety, the teacher identified one additional family at risk of injury in the home.

Validity:

1. Content validity: for the EC-Family Map, content validity was demonstrated in a sample of 26 Head Start centers with over 1,105 racially and ethnically diverse families by comparing the percentage of families at risk for a particular item to national estimates of similar risks (Whiteside-Mansell et al. 2007). For example, the EC-Family Map assessed about 60 percent of children living in poison-free homes, whereas the 2003 Head Start Family and Child Experiences Study estimated that 69 percent lived in these conditions. The IT-Family Map is closely based on the EC-Family Map for 3- to 5-year-olds. For the IT-Family Map, developers reviewed the literature, conducted pilot testing, and gathered input and consulted with Early Head Start and Head Start program staff and parents and experts.
2. Concurrent validity: the EC-Family Map discriminates between families who exhibit

homelessness, housing instability, and food insecurity; families with mothers experiencing depression; children with chronic illness or inadequate sleep; and parents with chronic illness (Whiteside-Mansell et al. 2010a). With regard to drug abuse, families who report household members' use of drugs and alcohol are at higher risk in all areas assessed by the EC-Family Map except for Nurturing Routines in the Routines subscale (Connors-Burrow et al. 2010; Bokony et al. 2010).

Bias Analysis: No information available.

Methods of Scoring: The assessor scores the Family MAP during the home visit. The simple scoring system is intended to be flexible such that assessors determine which modules are used during the home visit (there are no subscale or total scores). Interview and observation items include different types of response scales. For example, the parent may respond "yes," "no," or "don't know" when asked about an item or may provide a frequency for which the item occurs based on the response scale (e.g. "none" to "6 or more times" in the past week). Shaded responses on the form indicate the presence of a risk; non-shaded responses indicate no presence of a risk.

Interpretability: The form includes instructions to the assessor on how to interpret responses for items and subitems in the shaded range and whether the assessor should target the subscale containing the item for goal-setting. For example, within Environmental Safety, a set of questions probes for home and neighborhood exposure to injuries. If the parent indicates the

occurrence of any injury (responses are shaded), the assessor is instructed to consider targeting family goals to improve Environmental Safety outcomes.

Training Support: IT-Family Map training materials under development include video examples and instructional material in printable and web-based formats. A training manual and video demonstration of EC-Family Map administration during a home visit are available from the developer upon request. An eight-hour training session is recommended for all staff administering the interview; however, a four-hour training session is available for specialists with master's degrees in areas such as social work and early childhood development. Training topics emphasize the relationship between Head Start and families, how the home and parenting environment affect children's learning ability, interview techniques, and the development of goals based on outcomes of the interview.

Training costs include \$750 per half day for 20 participants in Little Rock, Arkansas. On-site training expenses also include the trainer's travel expenses. The developer may be contacted about costs for full-day training sessions and other arrangements.

Adaptations/Special Instructions for Individuals with Disabilities: The Family Map for the Hearing-Impaired Child is under development and will include an additional subscale that assesses how parents manage technological and medical needs and stress related to a child's hearing impairment.

Report Preparation Support: A summary score sheet is in development.

References:

Bokony, Patti A., Nicola A. Conners-Burrow, Leanne Whiteside-Mansell, Danya Johnson, Lorraine McKelvey, and Robert H. Bradley. "The Family Map: A Tool for Understanding the Risks for Children in Families with Substance Abuse." *NHSA Dialog: A Research-to-Practice Journal for the Early Intervention Field*, vol. 13, no. 3, 2010, pp. 192-197.

Conners-Burrow, Nicola A., Danya Johnson, Leanne Whiteside-Mansell, Lorraine McKelvey, Patti A. Bokony, and Robert H. Bradley. "Protecting Young Children: Identifying Family Substance Use and Risks in the Home." *NHSA Dialog: A Research-to-Practice Journal for the Early Intervention Field*, vol. 13, no. 3, 2010, pp. 168-184.

University of Arkansas for Medical Sciences. "Family Map of the Parenting Environment in Early Childhood." Available at [<http://familymedicine.uams.edu/FamilyMap/About>]. 2008.

Whiteside-Mansell, Leanne. *Family Map of the Parenting Environment in Early Childhood Instructional Manual*. Little Rock, AR: University of Arkansas Medical Sciences, 2008.

Whiteside-Mansell, Leanne, Patti A. Bokony, Robert H. Bradley, and Nicola A. Conners-Burrow. "Family Assessments to Enhance the Parent-Teacher Partnership." Paper presented to

the Arkansas Early Childhood Association. Hot Springs, AR, October 2008a.

Whiteside-Mansell, Leanne, Robert H. Bradley, Nicola A. Conners, and Patti A. Bokony. "The Family Map: Structured Interview to Identify Risks and Strengths in Head Start Families." *NHSA Dialog: A Research-to-Practice Journal for the Early Intervention Field*, vol. 10, 2007, pp. 189-209.

Whiteside-Mansell, Leanne, Robert H. Bradley, Nicola A. Conners-Burrow, Patti A. Bokony, Carol A. Lee, D.M. Mclemore, Danya Johnson, and L. Sockwell. "Documenting Head Start Impacts with the Family Map." Paper Symposium (Measurement Development)

presented to the Head Start Research Conference. Washington, DC, June 2008b.

Whiteside-Mansell, Leanne, Danya Johnson, Mary M. Aitken, Patti A. Bokony, Nicola A. Conners-Burrow, and Lorraine McKelvey. "Head Start and Unintended Injury: The Use of the Family Map Interview to Document Risk." *Early Childhood Education Journal*, vol. 28, 2010, pp. 33-41.

Whiteside-Mansell, Leanne, Danya Johnson, Nicola A. Burrow, Lorraine McKelvey, and Robert H. Bradley. "Evidence of Validity of the Early Childhood Family Map: A Tool to Assess Risk and Strength in Head Start." Poster presented to the Head Start Research Conference. Washington, DC, June 2010a.

FAMILY NEEDS SCALE (FNS)

Authors:

Carl J. Dunst, Carolyn S. Cooper, Janet C. Weeldreyer, Kathy D. Snyder, and Joyce H. Chase

Publisher:

Brookline Books
617-558-8010,
800-666-BOOK
www.brooklinebooks.com/

Initial Material Cost:

Book: Enabling and Empowering Families, \$25
Scales (10): \$10

Representativeness of Norming Sample:

No norming sample.

Languages:

English

Type of Assessment:

Parent self-report or report by other family member

Age Range and Administration Interval:

Families of young children

Personnel, Training, Administration, and Scoring Requirements:

Parent (or other family member) can complete the scale in 10 minutes. An early intervention practitioner can review the answers and interpret the scale, identifying places for concern, in under 10 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (.65 or higher) for internal and split-half reliability

Validity: 2 (< .5 for total score)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Family Needs Scale (FNS) is a 41-item scale that measures a family's needs in nine areas (financial, food and shelter, vocation, child care, transportation, communication, etc.). The parent (or family member) rates each item of need on a 5-point scale that ranges from (1) almost never a need to (5) almost always a need.

Uses of Information: The scale facilitates the identification of family needs and strengths. The results can be used to guide follow-up discussions to help clarify concerns and help define the precise nature of the family's needs. The information can then be used to decide whether intervention is required and, if so, the type of intervention needed. The book provides a number of case studies to illustrate how the information can be used.

Reliability: The research sample consisted of 54 parents of pre- and elementary school aged children who were mentally challenged, handicapped, and developmentally at risk.
(1) Internal reliability: coefficient alpha = .95;
(2) Split-half reliability (using the Spearman-Brown formula) = .96.

Validity: The authors tested for concurrent validity against a parent belief scale (Snyder et al. 1986). The FNS total scale score was found to be significantly related to the well-being (correlation = .42), decision-making (correlation = .40), and internal locus of control (correlation = .28) dimensions on the parent belief scale.

Methods of Scoring: The parent (or family member) reports the constancy of a need by marking Not Applicable, Almost Never, Seldom, Sometimes, Often, and Almost Always for each item of need.

Interpretability: Items rated Sometimes, Often, or Almost Always (a need) may indicate needs that are generally unmet, and thus provide a basis for further discussion to better understand the exact nature of the need. The book provides a Family Support Plan form and a Profile of Family Needs and Support form for the agency to use. The needs and support form can be used to record providers and the resources they are expected to provide to help the family address an identified need. The family support form enables the agency to mobilize resources to address needs and to monitor the progress of the intervention.

Training Support: None described. However, a second book published by Brookline Books, entitled “Supporting and Strengthening Families: Methods, Strategies, and Practices”(Dunst et al. 1994) is a collection of papers updating the

thinking and practices described in “Enabling and Empowering Families,” and building and elaborating upon the model described in the earlier book.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Dunst, Carl, Carol Trivette, and Angela Deal. *Enabling and Empowering Families: Principles & Guidelines for Practice*. Cambridge: Brookline Books, 1988.

Dunst, Carl, Carol Trivette, and Angela Deal. *Supporting and Strengthening Families: Methods, Strategies, and Practices*. Newton: Brookline Books, 1994.

Snyder, K.D., J.C. Weeldreyer, C.J. Dunst, and C.S. Cooper. *Parent Self-Awareness Scale: Reliability and Validity*. Unpublished scale. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1986.

FAMILY RESOURCE SCALE (FRS), 1986

Authors:

Carl J. Dunst and Hope E. Leet

Publisher:

Brookline Books

(617) 558-8010,

(800) 666-BOOK

www.brooklinebooks.com/

Initial Material Cost:

Book: Enabling and Empowering Families, \$25

\$10 per batch of 10 scales.

Representativeness of Norming Sample:

No norming sample.

Languages:

English

Type of Assessment:

Parent self-report or report by other family member

Age Range and Administration Interval:

Families of young children

Personnel, Training, Administration, and Scoring Requirements:

Parent (or other family member) can complete the scale in 10 minutes. An early intervention practitioner can review the answers and interpret the scale, identifying places for concern, in under 10 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (.65 or higher) for internal, split-half, and test-retest reliability

Validity: 3 (.5 or higher for criterion validity for total score)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The 31-item self-report Family Resource Scale (FRS) measures the adequacy of a family's tangible and intangible resources using a five-point scale, ranging from (1) not at all adequate to (5) almost always adequate. The scale covers such resources as food, shelter, financial resources, transportation, health care, time to be with family, child care, and time for self; which are generally organized from the most to the least essential resource. A modified version of the scales for teenage mothers is available.

Uses of Information: This scale determines the extent to which different types of resources are adequate in the households of young children. The lack of resources may be barriers to the family's involvement in their child's program, as families with unmet basic needs may not have

time or energy to participate actively in the child's program. The book provides a number of case studies to illustrate how the information can be used.

Reliability: The research sample consisted of 45 mothers of preschool retarded, handicapped, and developmentally at-risk children participating in an early intervention program. (1) Internal reliability (Cronbach's alpha): .92; split-half reliability (using the Spearman-Brown formula): .95 (2) Test-retest reliability (2 to 3 month interval): .52.

Validity: (1) Concurrent validity: both the personal well-being (Dunst 1986a) and maternal commitment (Dunst 1986b) measures were significantly related to the total scale score (.57 and .63, respectively).

Method of Scoring: The parent (or family member) marks the extent to which each of the resources is adequate for his/her family by selecting one of the following responses: Does Not Apply, Not At All Adequate, Seldom Adequate, Sometimes Adequate, Usually Adequate, and Almost Always Adequate.

Interpretability: Those items rated Not At All Adequate or Seldom Adequate may be evidence that these needs are not being met. They can provide a basis for exploring with the family the absence and need for these resources.

Training Support: None. However, a second book published by Brookline Books, entitled *Supporting and Strengthening Families: Methods, Strategies, and Practices* is a collection of papers updating the ideas and practices described in *Enabling and Empowering Families*, and building and elaborating upon the model described in the earlier book.

Adaptations/Special Instructions for Individuals with Disabilities: None

Report Preparation Support: The book provides a Family Support Plan form and a Profile of Family Needs and Support form for the agency

to use. The needs and support form can be used to record providers and the resources they are expected to provide to help the family address an identified need. The family support form enables the agency to mobilize resources to address needs and to monitor the progress of the intervention.

References:

Dunst, C.J. A Short Form Scale for Measuring Parental Health and Well-Being. Unpublished manuscript. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1986a.

Dunst, C.J. Measuring Parent Commitment to Professionally-Prescribed, Child-Level Interventions. Unpublished manuscript. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1986b.

Dunst, Carl, Carol Trivette, and Angela Deal. *Enabling and Empowering Families: Principles & Guidelines for Practice*. Cambridge: Brookline Books, 1988.

Dunst, Carl, Carol Trivette, and Angela Deal. *Supporting and Strengthening Families: Methods, Strategies, and Practices*. Newton: Brookline Books, 1994.

FAMILY SUPPORT SCALE (FSS), 1986

Authors:

Carl J. Dunst, Carol M. Trivette, and Vicki Jenkins

Publisher:

Brookline Books 617-558-8010

800-666-BOOK

www.brooklinebooks.com

Initial Material Cost:

Book: Enabling and Empowering Families, \$25

\$10 per batch of 10 scales

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Parent self-report

Age Range and Administration Interval:

Families of young children

Personnel, Training, Administration, and Scoring Requirements:

Parent can complete the scale in 10 minutes. An early intervention practitioner can review the answers and interpret the scale, identifying places for concern, in under 10 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 2 (.65 or higher for internal and split-half reliability; < .65 for test-retest reliability)

Validity: 2 (< .5 for criterion validity)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The self-report Family Support Scale (FSS) measures parents' satisfaction with the support they receive in raising a young child. The scale consists of 18 items covering such sources of support as the immediate family, relatives, friends and others in the family's social network, social organizations, and specialized and generic professional services. In addition, the scale provides 2 open items for parents to assess other sources of support not included in the 18 items. The parent rates each source of support on a 5-point Likert scale (ranging from not at all helpful (1) to extremely helpful (5)).

Uses of Information: Providers can use the FSS scale results to identify the areas in a family's support network that need to be strengthened or accessed to better meet the families' needs. The results can also be used to initiate inquiries into issues related to the support network. The FSS

might be useful as a pretest/posttest measure of perceived helpfulness of the program to the family (in relation to the family's level of involvement in the program). The book provides a number of case studies to illustrate how the information can be used.

Reliability: (1) Internal consistency reliability: coefficient alpha (on the 18-item scale) = .77; (2) Split-half reliability (using the Spearman-Brown formula): .75 (3) Test-retest reliability (1 month interval): correlation was .75 for the average correlation among the 18 scale items and .91 for the total scale scores. Test-retest reliability (18 month interval): correlation was .41 for the 18 scale items and .47 for the total scale scores.

Validity: (1) Criterion validity: The authors compared the results on the FSS scale to results on the Parent-Child Play Scale (Dunst 1986) and

selected subscales on the Questionnaire on Resources and Stress (Holroyd 1985). The FSS total scale score was consistently, but weakly, related to a number of parent and family outcomes, including personal well-being (correlation = .28), the integrity of the family unit (correlation = .18), parent perceptions of child behavior (correlation = .19), and opportunities to engage in parent-child play (correlation = .40) (Dunst 1985).

Method of Scoring: The parent answers how helpful various sources of support have been in terms of raising his/her child(ren) by circling Not Available, Not At All Helpful, Sometimes Helpful, Generally Helpful, Very Helpful, and Extremely Helpful.

Interpretability: A parent's responses are used to open up discussion as to why they use or do not use various means of support and resources. The book provides a Family Support Plan form and a Profile of Family Needs and Support form for the agency to use. The needs and support form can be used to record the names of providers and the resources they are expected to provide to help the family address an identified need. The family support form enables the agency to mobilize resources to address needs and to monitor the progress of the intervention.

Training Support: None described. However, a second book published by Brookline Books, entitled "Supporting & Strengthening Families: Methods Strategies and Practices" is a collection

of papers updating the thinking and practices described in "Enabling and Empowering Families," and building and elaborating upon the model described in the earlier book.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Dunst, C.J. "Rethinking Early Intervention." Analysis and Intervention Developmental Disabilities, vol. 5, 1985, pp. 165-201.

Dunst, C.J. A Rating Scale for Assessing Parent-Child Play Opportunities. Unpublished scale. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1986.

Dunst, Carl, Carol Trivette, and Angela Deal. Enabling and Empowering Families: Principles & Guidelines for Practice. Cambridge: Brookline Books, 1988.

Dunst, Carl, Carol Trivette, and Angela Deal. Supporting and Strengthening Families: Methods, Strategies, and Practices. Newton: Brookline Books, 1994.

Holroyd, J. Questionnaire on Resources and Stress Manual. Unpublished scale. Los Angeles: University of California, Neuropsychiatric Institute, Department of Psychiatric and Behavioral Sciences, 1985.

THE HOME OBSERVATION FOR MEASUREMENT OF THE ENVIRONMENT INVENTORY FOR INFANTS/TODDLERS (IT-HOME) AND EARLY CHILDHOOD (EC-HOME), 2003 (2011 Update)

Authors:

Bettye M. Caldwell and Robert H. Bradley

Publisher:

University of Arkansas

<http://ualr.edu/case/index.php/home/home-inventory/>
(501) 565-7627

Initial Material Cost:

Standard Manual \$40, Infant Toddler forms \$15 per pad, Early Childhood forms \$25 per package of 50, Comprehensive Manual with information on all the inventories \$50

Materials for administering the HOME Inventory available by contacting lrcoulson@ualr.edu

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Home observation and parent interview

Age Range and Administration Interval:

Birth to age 3 years. The Early Childhood HOME Inventory is for children 3 to 6 years old. It may be administered at desired intervals.

Personnel, Training, Administration, and Scoring

Requirements: Formal training is not required but is recommended. Training DVDs, interviews for use in training and recalibration, and scoring documentation entitled "Scoring Keys and Rationale" are available from Cooperative Extension Publications at <http://learningstore.uwex.edu> for \$25.

The HOME Inventories take between 45 and 90 minutes to administer.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability:¹ 3 (all .65 and higher for inter-rater and internal consistency), 2 (mostly under .65 for test-retest)

Validity:¹ 2 (less than .5 for concurrent), 3 (mostly .4 or higher for predictive)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 2 (administered by a highly trained individual; scored by someone with basic clerical skills)

¹ Reliability and validity information for the IT- and EC-HOME are based on the HOME Inventory Administration Manual (2001) and other studies of previous versions of the HOME. The 2003 manual provided no updated information.

Description: The Home Observation for Measurement of the Environment (HOME) Inventory is designed to measure the quality and extent of stimulation available to a child in the home environment. The Infant/Toddler HOME Inventory (IT-HOME) comprises 45 items that provide information from the child's perspective on stimuli found to affect children's cognitive development. Assessors make observations during home visits when the child is awake and engaged in activities typical for that time of the day and

conduct an interview with a parent or guardian. The IT-HOME is organized into six subscales: (1) Responsivity: the extent of responsiveness of the parent to the child; (2) Acceptance: parental acceptance of suboptimal behavior and avoidance of restriction and punishment; (3) Organization: including regularity and predictability of the environment; (4) Learning Materials: provision of appropriate play and learning materials; (5) Involvement: extent of parental involvement; and (6) Variety in daily stimulation. For the IT-

HOME, 18 items are based on observation, 15 on interview, and 12 on either observation or interview.

Separate inventories are available for early childhood (EC-HOME, age 3 to 6 years), middle childhood (MC-HOME, age 6 to 10 years), and early adolescence (EA-HOME, age 10 to 15 years).² The IT-, EC-, and MC-HOME versions have also been adapted for children with disabilities. The EC-HOME includes 55 items organized into eight subscales: (1) Learning Materials; (2) Language Stimulation: overt attempts to foment language development; (3) Physical Environment; (4) Responsivity of Parent to Child; (5) Academic Stimulation: parental involvement in child's intellectual development; (6) Modeling: parents' demonstration of desirable behaviors; (7) Variety in Daily Stimulation and Enrichment; and (8) Acceptance: parental acceptance of suboptimal behavior and avoidance of unnecessarily harsh restriction or punishment (same as subscale 2 on the IT-HOME). Twenty-one items are based on observation, and 12 may be based on observation or interview. As with the IT-HOME, the EC-HOME is administered by a visitor to the family's home at a pre-scheduled time when the child is awake and engaged in activities typical for that time of the day. The 2003 IT- and EC-HOME were updated from the 2001 versions, slightly re-ordering questions within subscales and including a structured interview

² This summary focuses on the IT-HOME and EC-HOME. While their items and subscale topics differ, the assessments for all age groups share a similar structure and involve similar administration and scoring protocols.

format (versus informal administration of questions) and a training DVD.

Child Care versions of the IT-HOME and EC-HOME are also available. The Child Care HOME Inventory evaluates a family child care environment in a home setting, but adult caregivers such as babysitters, nannies, or daycare providers are the object of interest rather than parents or guardians. Items thus refer to a "caregiver" instead of a "parent." The content of the assessment is otherwise identical.

The Supplement to the HOME for Impoverished Families (SHIF) assesses the quality of the home environment of young children living in poor urban homes. The SHIF is a supplemental set of questions specific to low-income homes and should be used with the total HOME Inventory rather than as an independent assessment. The 20-item SHIF takes approximately eight minutes to score and uses the same scoring procedures as the HOME Inventory.

Other Languages: None.

Uses of Information: The HOME Inventory identifies environments that do not stimulate the cognitive development of children and informs the development of interventions that benefit both caregiver and child. The HOME Inventory also measures change in both the family environment and the quality of parenting when interventions are implemented (Totsika and Sylva 2004).

Reliability:

1. Internal consistency reliability: Cronbach's alphas were .84 for the IT-HOME and ranged

from .49 to .78 for the six subscales.³ Kuder-Richardson coefficients were .89 for the inventory and ranged from .44 to .89 for the subscales. For a non-representative sample of 73 poor urban families with a child age 3 years or younger, Kuder-Richardson coefficients obtained for the HOME Inventory and SHIF, respectively, were .80 and .63 (Ertem et al. 1997).

2. Test-retest reliability: Saudino and Plomin (1997) reported a coefficient of .94 on a normative sample at 12 months (with a two-week interval). They also reported stability between 12 and 24 months at .64. Pearson and intraclass correlations measured the stability in IT-HOME scores between administrations overall and by subscale and by age group.
3. Inter-rater reliability: the Kappa statistics for inter-rater reliability between teams of pediatricians and research assistants ranged from .76 to 1.0 for the HOME and .79 to 1.0 for the SHIF.

Validity:¹

1. Concurrent validity: in a non-representative sample of 73 impoverished urban families, the Pearson correlation between the HOME Inventory and the SHIF was .69 and .55 and .42 between the HOME Inventory and the Nursing Child Assessment Feeding Scale (NCAFS) and the Nursing Child Assessment

Teaching Scale (NCATS), respectively, and .49 and .36 between the SHIF and the NCAFS and the NCATS (Ertem et al. 1997).

The IT-HOME detects differences in home environments characterized by poverty and/or maternal learning disabilities (Keltner 1994), maternal responsiveness and provision of play stimulation subscales for mothers with psychiatric disorders such as schizophrenia and depression (Goodman and Brumley 1990), and aspects of the home potentially detrimental to medically fragile infants (Holditch-Davis et al. 2000). The IT-HOME does not detect differences in the cognitive development scores of infants with substance-abusing mothers and those with healthy mothers, although it is possible that the HOME scoring system is not sufficiently sensitive to pick up on individual differences in some samples of mothers (Howard et al. 1995; Beckwith 1996).

2. Predictive validity: IT-HOME scores administered at 6, 12, and 24 months were compared to the child's scores on the Bayley Scales of Infant Development MDI at 12 months, the Stanford-Binet at 36 and 54 months, and the ITPA at 37 months. The IT-HOME was a better predictor of intelligence than the socioeconomic measures and was a stronger predictor for females and whites. The IT-HOME for low-birth weight and pre-term babies is an independent predictor of cognitive development among children age 2 years and older (Weisglas-Kuperus et al. 1993;

³ The analysis was performed on the 45 scale items that were part of a longer (72 items) version administered to a sample of 232 families in Syracuse, New York.

Molfese et al. 1996). HOME Inventory scores obtained for children age 2 years and older were more predictive of later cognitive development when correlated with mental measures than scores obtained earlier for infants 6 or 12 months old (Elardo et al. 1975; Bradley and Caldwell 1976; Bradley and Caldwell 1979; Bradley, Caldwell, and Rock 1988). Findings from a National Institute of Child Health and Human Development (NICHD) study revealed that the IT- and EC-HOME measures of maternal responsiveness and sensitivity (along with child gender and family income) were strong predictors of children's later attachment patterns (NICHD Early Child Care Research Network 2001).

Bias Analysis: Bradley et al. (1994) conducted a factor analysis to determine the suitability of the IT- and EC-HOME inventories for different ethnic groups. Using a sample of families of pre-term infants, they concluded that the factor structures were in agreement with the organization of the items into subscales for the black and white samples, but less so for the Hispanic sample. Bradley et al. (1989) conducted a large longitudinal study on the relationship between HOME Inventory scores, socioeconomic status, and cognitive measures for three ethnic groups: whites, blacks, and Mexican Americans. Scores were well correlated with cognitive development for the white group, less so for the black group, and almost not at all for the Mexican American group.

Methods of Scoring: The assessor enters a plus sign for each item if the behavior is observed or reported and a minus sign if it is not. Subscale and total inventory scores are derived by counting the number of plus signs. Scores are given at the time of observation, with no interpretation required after the visit.

Interpretability: The summary sheet provides the scores that fall into the lowest quartile, the middle half, and the upper quartile. Homes with scores in the lowest quartile are considered environments that pose an increased risk to children's development. However, the authors recommend that the interviewer take notes while chatting with the parent and/or record information gleaned from the ice-breaker questions on the child and the family in order to provide a context for interpreting the HOME Inventory scores. In addition, the assessor should pay attention to patterns across the subscales; such patterns may provide information for use in developing and structuring an intervention. For the sample of 73 poor urban families, the mean SHIF score was 15.5 with a standard deviation of 2.8, although no information was provided on interpreting SHIF scores.

Training Support: The authors and trainers offer workshops, and training DVDs are available from Cooperative Extension Publications at <http://learningstore.uwex.edu>. Information on the workshop and the DVDs is available from the authors.

Adaptations/Special Instructions for Individuals with Disabilities: The IT-, EC-, and

MC-HOME versions have been adapted for children with disabilities. Versions of the Disability-adapted HOME are available for the three age groups for four categories of disabilities: Developmental Delay, Auditory Impairment, Visual Impairment, and Orthopedic Impairment. Separate instructions and new item definitions within each subsection accompany each version.

Report Preparation Support: A summary sheet for recording background information on the family, subscale, and total inventory scores is available.

References:

- Beckwith, L. "Measurement of the Early Rearing Environment: Caregiver-Child Interaction." Report No. 166. Citeseer, 1996.
- Bradley, R.H., D.J. Mundfrom, L. Whiteside, P.H. Casey, and K. Barrett. "A Factor Analytic Study of the Infant-Toddler and Early Childhood Versions of the HOME Inventory Administered to White, Black, and Hispanic American Parents of Children Born Preterm." *Child Development*, vol. 65, no. 3, 1994, pp. 880-888.
- Caldwell, Bettye M., and Robert H. Bradley. *HOME Inventory Administration Manual*. Little Rock, AR: University of Arkansas for Medical Sciences and University of Arkansas at Little Rock, 2001.
- Caldwell, Bettye M., and Robert H. Bradley. *HOME Inventory Administration Manual, Comprehensive Edition*. Little Rock, AR: University of Arkansas for Medical Sciences and University of Arkansas at Little Rock, 2003.
- Caldwell, Bettye M., and Robert H. Bradley. *Home Observation for Measurement of the Environment (HOME - Revised Edition)*. Little Rock, AR: University of Arkansas, 1984.
- Ertem, Ilgi O., Brian W.C. Forsyth, Abraham J. Avni-Singer, Lisa K. Damour, and Dominic V. Cicchetti. "Development of a Supplement to the HOME Scale for Children Living in Impoverished Urban Environments." *Developmental and Behavioral Pediatrics*, vol. 18, no. 5, 1997, pp. 322-328.
- Goodman, S.H., and H.E. Brumley. "Schizophrenic and Depressed Mothers: Relational Deficits in Parenting." *Developmental Psychology*, vol. 26, no.1, 1990, pp. 31-39.
- Holditch-Davis, D., E.M. Tesh, B.D. Goldman, M.S. Miles, and J. D'Auria. "Use of the HOME Inventory with Medically Fragile Infants." *Children's Health Care*, vol. 29, 2000, pp. 257-278.
- Howard, J., L. Beckwith, M. Espinosa, and R. Tyller. "Development of Infants Born to Cocaine-Abusing Women: Biologic/Maternal Influences." *Neurotoxicology and Teratology*, vol. 17, no. 4, 1995, pp. 403-411.
- Keltner, B. "Home Environments of Mothers with Mental Retardation." *Mental Retardation*, vol. 32, no. 2, 1994, pp. 123-127.
- Molfese, J.V., L.F. DiLalla, and L. Lovelace. "Perinatal, Home Environment and Infant Measures as Successful Predictors of Preschool Cognitive and Verbal Abilities." *International Journal of Behavioral Development*, vol. 19, no. 1, 1996, pp. 101-119.

Saudino, J.K., and R. Plomin. "Cognitive and Temperamental Mediators of Genetic Contributions to the Home Environment during Infancy." *Merrill-Palmer Quarterly*, vol. 43, no. 1, 1997, pp. 1-23.

Totsika, Vasiliki, and Kathy Sylva. "The Home Observation for Measurement of the Environment Revisited." *Child and Adolescent Mental Health*, vol. 9, no. 1, 2004, pp. 25-35.

Weisglas-Kuperus, N., W. Baerts, M. Smrkovsky, and P.J.J. Sauer. "Effects of Biological and Social Factors on the Cognitive Development of Very Low Birth Weight Children." *Pediatrics*, vol. 92, no. 5, 1993, pp. 658-665.

Weisglas-Kuperus, N., H.M. Koot, W. Baerts, W.P.F. Fetter, and P.J.J. Sauer. "Behaviour Problems of Very Low Birth Weight Children." *Developmental Medicine and Child Neurology*, vol. 35, no. 5, 1993, pp. 406-416.

INFANT-TODDLER AND FAMILY INSTRUMENT (ITFI), 2001

Authors:

Nancy H. Apfel and Sally Provence

Publisher:

Paul H. Brookes Publishing Co.

(800) 638-3775

www.brookespublishing.com

Initial Material Cost:

ITFI instrument and manual package: \$45

Representativeness of Norming Sample:

Not normed. Field test involved 55 Connecticut families with 59 children ages 6 to 36 months.

Languages:

English

Type of Assessment:

Direct parent and child assessment; structured child observation (parent report if observation not possible)

Age Range and Administration Interval:

6 months to 3 years

Personnel, Training, Administration, and Scoring

Requirements: Designed for home visitors. Can also be used by family practitioners, including paraprofessionals, who have varying levels of training. Preparation sessions for orientation and instructional purposes are strongly recommended.

The Caregiver Interview and Developmental Map can be administered in two 45- to 60-minute sessions. One 45- to 60-minute session is needed to share findings and develop a follow-up plan for the family.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by family service workers)

Description: The ITFI helps family service providers assess the well-being of children 6 months to 3 years of age and their families. It consists of four sections: the Caregiver Interview, Developmental Map, Checklist for Evaluating Concern, and the Plan for the Child and Family. The Caregiver Interview has 35 items related to home and family life, child health and safety, and family issues and concerns. The interview is conducted with the parent(s) or primary caregiver(s). The Developmental Map is an observation of infant-toddler behavior in the areas of gross and fine motor development, social and emotional development, language development, and coping and self-help development. It involves informally observing the child's behavior and

interaction with others, as well as playing with the child using an established set of testing materials. The Checklist for Evaluating Concern is completed by the interviewer after the visit in order to assess the family in the areas of home and family environment; child health, development and safety; and stressors in the child's life. The interviewer rates concerns on a scale of 1 (low) to 10 (high) and prioritizes concerns based on these ratings. The Plan for the Child and Family determines what steps may need to be taken in order to address the interviewer's concerns about the needs of the family.

Uses of Information: Family service providers can use the ITFI as a supplement to other assessments of child and family strengths

and needs, specific child symptoms and stressors, and the caregiver's ability to meet their child's basic needs. It also helps service providers work with families to develop a support plan for meeting their needs.

Reliability: None described.

Validity: None described.

Method of Scoring: Scoring is done on the three-part Checklist for Evaluating Concern, after the family service provider leaves the family's home. The checklist summarizes the family provider's impressions of family and child needs and strengths based on information from the Caregiver Interview, the Developmental Map, and observations of the caregiver-child interaction and the home environment. For each item in the checklist, the provider indicates whether the condition is present, is of concern, or if the provider is unsure of its presence.

Interpretability: The interviewer uses the Checklist for Evaluating Concern Summary Sheet to rate the level of concern for the child and family from 1 (low) to 10 (high) and to list the family's strengths and weaknesses. The ratings of concern from the Checklist for Evaluating Concern may determine how detailed a support plan is, when it is put into action, and how intense services should be. After the service provider has considered all of

the information from the summary sheet, he or she is able to prioritize the concerns while keeping the child's health, development, and safety as a top priority. The manual provides general guidelines and case studies on how to use information collected with the ITFI that can be used to prioritize needs and develop service plans. The guidelines are kept general to allow agencies and programs to incorporate the ITFI into their own protocol, purpose, and service.

Training Support: "Brookes on Location" professional development seminar, Using ITFI to Evaluate Young Children and Their Families, is available through the publisher.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: Case study examples are provided in which a case is presented, family strengths and vulnerabilities are highlighted, and an action plan for intervention is developed.

References:

Apfel, Nancy and Sally Provence. Manual for the Infant-Toddler and Family Instrument (ITFI).

Baltimore: Paul H. Brookes Publishing Co., 2001.

THE KEMPE FAMILY STRESS INVENTORY (KFSI)

Authors:

Barton Schmitt and Claudia Carroll with assistance from Jane Gray

Publisher:

Authors

<http://www.ucdenver.edu/academics/colleges/medicalschool/departments/pediatrics/subs/can/Pages/ChildAbuseNeglect.aspx>

Initial Material Cost:

Free. The KFSI is not copyrighted and can be used at no charge and without the authors' permission, but permission is needed to re-publish the measure; supplemental rating criteria are copyrighted.

Representativeness of Norming Sample:

None described.

Languages:

English

Type of Assessment:

Direct Parent Assessment

Age Range and Administration Interval:

Not applicable. Focuses on parents of all ages.

Personnel, Training, Administration, and Scoring

Requirements: Personnel need to have appropriate training or experience to conduct psychosocial interview, and must have specialized training to use supplemental rating criteria.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: Inter-rater: 3 (.65 or higher)

Validity: 2 (< .4 for predictive)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual).

Description: The KFSI assesses parents' risk for child maltreatment and/or caregiving difficulties. It is a 10-item rating scale that is completed after a through psychosocial interview with a trained professional. Service providers, such as home visitors, may also rate families based on the interactions they have had with them over a period of time (Korfmacher, Younge, and Michalek 1996). Items on the KFSI assess parents on a number of domains, such as psychiatric and criminal history, childhood history of care, emotional functioning, attitudes towards and perception of children, discipline of children, and level of stress in the parent's life. Parents receive a raw score and are determined to be at low, moderate, or high risk, depending on the cut-offs

established by the program administering the scale (see section on interpretability, below).

Uses of Information: The KFSI is currently used with at-risk families as an integral part of the screening and assessment process for two home-visiting programs for families at-risk, the Hawaii Healthy Start and the nation-wide Healthy Families America program. It is primarily used as a second-level screening tool for mothers who are considered at risk based on a 15-item hospital chart review that assesses sociodemographic risk factors, such as maternal age and income.

Reliability: (1) Inter-rater reliability: a Healthy Families America program site in Oregon collected data on 115 families, and notes from each of their psychosocial interviews were reviewed by independent raters. The reliability

correlation coefficient between raters on classifying parents into the low, mild, or severe risk categories was .93. (Katzev et al. 1997).

Validity: Predictive validity: three studies that compared KFSI scores to child maltreatment based on hospital records, state child protection reports, and scores on the Child Abuse Potential Inventory have shown KFSI to have predictive validity (Murphy et al. 1985; Hawaii Department of Health 1992; and Katzev et al. 1997). Three studies of predictive validity were conducted using varying cut-off scores and time periods. Positive predictive validity reflected the percentage of mothers with high-risk scores who later maltreated their children, and ranged from 25 to 52 in two of the studies. According to Korfmacher (2000), the low percentage (3 percent) in a third study may be an artifact of the study's reliance on child protection reports to measure the incidence of maltreatment. Negative predictive validity measured the percentage of mothers with low-risk scores who demonstrated no future evidence of child maltreatment; it ranged from 85 to 100 percent. The KFSI sensitivity, which is the percentage of maltreating mothers who were scored at high risk, ranged from 80 percent to 97 percent across the three studies. The specificity, which is the percentage of non-maltreating mothers who were scored at low risk, ranged from 21 to 89 percent.

Method of Scoring: A three-point scale, ranging from low to high risk, is used to score both parents. Parents who receive a high-risk score are considered at risk for child

maltreatment. Two methods have been employed in scoring the KFSI. Carroll (1978) used weighted scoring, in which items that were determined to be more immediate precedents to child abuse (such as violent outbursts and harsh punishment of child) were given higher scores, while lower weighting was given to items that were perceived to be less important in assessing immediate risk (such as parents' history of child maltreatment, low self-esteem or isolation, and having an unwanted child). If a weighted scaling system is not used, items are assigned values according to whether there is no risk (0), risk (5), or high risk (10) of child maltreatment. The total score is obtained by summing the values assigned to each item, although the weighted system requires that examiners multiply raw scores by 2.5. Total scores range from 0 to 100.

Interpretability: KFSI users have applied different cutoff scores based on their clinical judgment to identify at-risk parents. Carroll (1978) felt that scores above 30 were "concerning" and that scores above 70 were "particularly differentiating"; however, others have defined different cutoffs and have assigned labels (such as, low, medium and severe risk) to scores in a particular range. To aid in interpreting scores, Healthy Families America and the Family Stress Center have created and copyrighted supplemental rating criteria.

Training Support: Individuals must participate in a specialized training on using the supplemental rating criteria, which have been

copyrighted by Healthy Families America and the Family Stress Center.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Carroll, C.A. "The social worker's evaluation." In *The child protection team handbook*, edited by B.D. Schmitt. New York: Garland STM Press, 1978, pp. 83-108.

Center on Child Abuse Prevention Research. *Intensive home visitation: A randomized trial, follow-up, and risk assessment study of Hawaii's Healthy Start program*. Final report prepared for the National Center on Child Abuse and Neglect. Chicago, IL: National Committee to Prevent Child Abuse, 1996.

Hawaii Department of Health. Report to the 16th Legislature, State of Hawaii, on House Bill #139, c.d. 1: Requesting review and

recommendations from the Director of Health on the Healthy Start Program. Honolulu, HI: Maternal and Child Health Branch, Hawaii Department of Health, 1992.

Katzev, A., T. Henderson, and C. Pratt. Predicting child maltreatment with the Kempe Family Stress Assessment. Document presented at HFA: Rethinking the Assessment Process working meeting. Chicago, IL: National Committee to Prevent Child Abuse, 1997.

Korfmacher, J., P. Younge, and P. Michalek. Standardized assessment of child abuse risk: A review. Paper presented at 24th Annual Child Abuse and Neglect Symposium, Keystone, CO, 1996.

Korfmacher, Jon. The Kempe Family Stress Inventory: A review. *Child Abuse and Neglect*, vol. 24, 2000, pp. 129-140.

Murphy, S., B. Orkow, and R.M. Nicola. Prenatal prediction of child abuse and neglect: A prospective study. *Child Abuse & Neglect*, vol. 9, 1985, pp. 225-235.

KNOWLEDGE OF INFANT DEVELOPMENT INVENTORY (KIDI), 1981

Authors:

David MacPhee

Publisher:

Unpublished manuscript; available from Educational Testing Service

(609) 734-5689

www.ets.org/

Initial Material Cost:

As of January 1998, the cost was \$11, plus \$3 shipping and handling to order this measure from the Educational Testing Service

Representativeness of Norming Sample:

Non-representative sample of pediatricians, Ph.D.s in child psychology, University of North Carolina undergraduate child psychology students, and mothers in Chapel Hill, NC.

Languages:

English

Type of Assessment:

Parent report

Age Range and Administration Interval:

Parents of infants and young children

Personnel, Training, Administration, and Scoring

Requirements: An individual with a 7th-grade reading level can complete the instrument. Administration time is 20 minutes; scoring time is 20 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 2 (some less than .65; others .65 or higher)

Validity: 1 (validity coefficients not reported)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 2 (self-administered; scored by someone with basic clerical skills)

Description: The Knowledge of Infant Development Inventory (KIDI) is a 75-item instrument that was designed to obtain comprehensive information on parents' factual knowledge of parental practices, child developmental processes, and infant norms of behavior. The KIDI is designed to be easily accessible to persons with limited education and to be culturally neutral. The items can also be grouped into four non-exclusive general categories to obtain more specific information on a person's knowledge on infant norms and milestones, principles of infant development, parenting, and health and safety. The KIDI Scale is accompanied by a 17-item questionnaire (the Catalog of Previous Experience, or COPE)

assessing previous experience with infants to correlate with knowledge level assessed by KIDI.

Uses of Information: The KIDI may be used as an indicator or a diagnostic tool for high-risk parents and also to evaluate parent education programs.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): alphas were .67 and .55 for college students at pretest and posttest, respectively, .82 for parents, and .50 for professionals. The Guttman split-half coefficients were .60 and .57 for college students at pretest and posttest, respectively, .85 for mothers, and .59 for professionals. (2) Test-retest reliability: For parents (2-week interval), the correlation coefficients were .92 for the total score, .80 for

attempted, and .91 for accuracy. For college students (4-month interval), the coefficients were .65 for the attempted and .47 for accuracy.

Validity: (1) Content validity: The author conducted an extensive review of the relevant literature and the instrument has been reviewed by parents, pediatricians, and persons holding a Ph.D. in child psychology. (2) Construct validity: The manual reports the results of the initial validity studies conducted by the author. The results suggested that persons with more experience with or knowledge about infants were more confident in responding to the KIDI. However, persons with formal knowledge were more accurate in their responses than persons with informal knowledge. (3) Predictive validity: In another study, the author found parents of developmentally delayed children had significantly lower KIDI attempted and accuracy scores than parents of children with normal development.

Method of Scoring: Each of the KIDI items is scored as right (+1), wrong (-1), or not sure (0) according to an answer key that is provided. Using formulas provided in the manual, three summary scores are then calculated: an attempted score (percent of items attempted, a measure of confidence), an accuracy score (percent correct of the attempted answers), and a total correct score (percent correct of all the KIDI items). If the user

wishes, subscale scores can be calculated for the four general categories: (1) norms and milestone, (2) principles, (3) parenting, and (4) health and safety.

Interpretability: No instructions provided.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

MacPhee, D. Manual: Knowledge of Infant Development Inventory. Unpublished manuscript, University of North Carolina, 1981.

MacPhee, D. The Nature of Parents' Experiences with and Knowledge About Infant Development. Paper presented at the Society for Research in Child Development, April 1983.

MacPhee, D. The Relationship Between Children's Delayed Development and Their Mothers' Perceptions of Development. Paper presented at the Southwestern Society for Research in Human Development, March 1984.

MacPhee, D. Mothers' Acquisition and Reconstruction of Knowledge About Infancy. Paper presented at the Southwestern Society for Research in Human Development, March 1984.

NURSING CHILD ASSESSMENT SATELLITE TRAINING (NCAST) PARENT-CHILD INTERACTION PROGRAM

NURSING CHILD ASSESSMENT TEACHING SCALE (NCATS), 2ND EDITION, 1995

Authors:
NCAST

Publisher:
NCAST
(206) 543-8528
www.ncast.org

Initial Material Cost:
Teaching set (includes teaching manual, scale pad, and teaching kit): \$125

Representativeness of Norming Sample:
Diverse but non-random sample. Sample consists of approximately 2,100 observations sent in by persons across the United States learning to use the scale from around 1980 to 1995.

Languages:
English

Type of Assessment:
Observation

Age Range and Administration Interval:
Birth to 36 months

Personnel, Training, Administration, and Scoring Requirements:
Administered by a professional health care worker certified by NCAST as a learner or instructor. Usually administered in 1 to 6 minutes.

Summary:
Initial Material Cost: 2 (\$100 to \$200)
Reliability: 3 (some subscales fell below .65, although most of the total scales exceeded .65)
Validity: 2 (< .5 for concurrent, < .4 for predictive)
Norming Sample Characteristics: 2 (not nationally representative)
Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Nursing Child Assessment Teaching Scale (NCATS) is used to assess the quality of the caregiver-child teaching interaction for children from birth to 3 years of age. The 73-item teaching scale is organized into six subscales, four of which assess the caregiver's behavior and two the child's. The four caregiver subscales assess the caregiver's sensitivity to cues, response to the child's distress, fostering of social-emotional growth, and fostering of cognitive growth. The two child subscales assess the clarity of the child's cues and responsiveness to the caregiver.

Uses of Information: The teaching scale identifies areas of strengths and weaknesses in the caregiver-child teaching interaction. The results can be used to build the caregiver's skills to facilitate the development of the child.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): ranged from .52 to .80 on the caregiver subscales, .50 on the child's clarity of cues, and .78 on the child's responsiveness to parent subscales. The alphas for the total caregiver and child subscales were .87 and .81, respectively. (2) Test-retest reliability (with a 3- to 4-month interval between tests): .85 on the total parent score and .55 on the total infant score.

Validity: (1) Concurrent validity: NCATS caregiver scores were tested for concurrent validity against the Home Observation for Measurement of the Environment (HOME) and the Bayley Scales of Infant Development. The correlations of the total NCATS scores with the total HOME score among children ages 1 to 36 months, in three age groups, ranged from .41 to

.44. The correlation of the total NCATS score with the Bayley Mental Development Index (MDI) and Bayley Psychomotor Development Index (PDI) were .28 and .34, respectively. In both cases, the caregiver scales, especially the social-emotional and cognitive growth subscales, were more strongly correlated with the HOME and Bayley. (2) Predictive validity: a test for predictive validity reported correlations of .23 and .34, respectively, between NCATS total scores taken at 3 and 10 months and MDI scores, both statistically significant. The subscale correlations ranged from -.01 to .37. Correlations between the NCATS caregiver and total scores at 24 months with the Bayley MDI (at 24 months), Preschool Language (at 36 months), and WPPSI IQ (at 60 months) were stronger and more consistent than the correlations between the cognitive measures and NCATS scores at 12 months.

Method of Scoring: During the teaching session, the observer goes through the 73-item scale and marks “yes” or “no” for each item depending upon whether or not the behavior was observed. The teaching manual provides the user with step-by-step scoring instructions. The user must calculate the totals for each subscale and the total score. The scores are compared to a table provided in the manual to determine whether the score falls under the 10 percentile cutoff score.

Interpretability: The Teaching Manual provides step-by-step instructions on how to

interpret scores using population norms. Tables that compare NCATS scores to the norms are provided.

Training Support: To learn essential child care skills, users of NCATS are strongly recommended to view NCAST’s “Keys to Caregiving” video series. Workshops are also available through NCAST or NCAST certified instructors. The fee for NCAST’s workshops is \$900, which covers training on assessing caregiver-child interaction. NCAST recommends that individuals view the “Keys to Caregiving” videos prior to attending NCAST training.

Adaptations/Special Instructions for Individuals with Disabilities: None.

Report Preparation Support: General guidance is provided in the step-by-step instructions on the use of the teaching scale, including some suggestions on how to review and discuss any identified problems with caregivers and how to document an agreed upon prescription of practice/behavior by the caregivers.

References:

Summer, Georgina and Anita L. Spietz. NCAST Caregiver/Parent-Child Interaction Teaching Manual, 2nd Edition. Seattle, WA: NCAST Publications, University of Washington, June 1995.

PARENTAL MODERNITY SCALE, 1985 (2011 Update)

Authors:

Earl S. Schaefer and Marianne Edgerton

Publisher:

None

Initial Material Cost:

None

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Parent report

Age Range and Administration Interval:

Parents of infants, toddlers, and elementary school children

Personnel, Training, Administration, and Scoring Requirements: No information available**Summary:**

Initial Material Cost: 1 (< \$100)

Reliability: 3 (mostly .65 or higher)

Validity: 3 (all .5 or higher for concurrent)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 2 (self-administered and/or administered and scored by someone with basic clerical skills)

Description: The Parental Modernity Scale (Schaefer and Edgerton 1985) is a 30-item measure of parental beliefs that are classified as traditional and authoritarian or progressive and democratic. The scale was designed for use with parents of school-age children. Parents indicate whether they strongly disagree, mildly disagree, are not sure, mildly agree, or strongly agree with each item statement. Alternatively, the assessor may read the statements aloud and prompt the parent for a response. Eight items belong to the Progressive Beliefs subscale (6, 11, 13, 15, 20, 23, 27, and 29), and the remaining 22 items belong to the Traditional Beliefs subscale.

The Parental Modernity Scale was adapted for use with parents of toddlers in the Early Head Start Family and Child Experiences Survey (Baby FACES) and the Early Head Start Research and Evaluation Project (EHSREP). Ten items out of 30 were selected for the two studies and then divided evenly between the two subscales. Traditional

Beliefs included items 3, 5, 10, 21, and 28, and Progressive Beliefs included items 6, 13, 20, 23, and 29.

Other Languages: None.

Uses of Information: The Parental Modernity Scale and its adapted shorter form have been used to conduct longitudinal research on child care experiences in the National Institute of Child Health and Human Development's Study of Early Child Care and Youth Development (NICHD SECCYD) in order to measure traditional and progressive parenting beliefs. Both forms have also been used in the Baby FACES study and the EHSREP.

Reliability:

1. Internal consistency reliability: research studies have published Cronbach's alpha coefficients for the Parental Modernity Scale subscales. The Progressive Beliefs and Traditional Beliefs subscale coefficients were

.68 and .73 (Administration for Children and Families 2002). Vogel and colleagues (2011) reported Cronbach's alpha coefficients of .59 and .58 for the adapted Traditional Beliefs and Progressive Beliefs subscales, respectively.

2. Test-retest reliability: no information available.
3. Inter-rater reliability: not applicable.

Validity:

1. Content validity: the authors correlated items about parental beliefs from a scale they had developed (the Parent as Educator Interview) with measures of kindergarten children's academic competence. They selected the 30 items significantly correlated with academic competence, of which more items were related to traditional, authoritarian beliefs because the items were more highly correlated with children's academic competence.
2. Concurrent validity: in a sample of 49 parents, the authors reported total parental modernity scores correlated with teacher ratings of child verbal intelligence at .50 and .51 for mothers and fathers, respectively, and at .55 combined.

Bias Analysis: No information available.

Methods of Scoring: Items on the Parental Modernity Scale are rated on a five-point scale ranging from 1 = strongly disagree to 5 = strongly agree. The Progressive Beliefs subscale score is the sum of 8 items (6, 11, 13, 15, 20, 23, 27, 29), and the Traditional Beliefs subscale score is the sum of the 22 remaining items. The NICHD SECCYD

researchers also computed a Total Traditional Beliefs score, which included the sum of the Traditional items and the reverse scores of Progressive items.

Interpretability: Higher Progressive Beliefs scores reflect modern beliefs about childrearing and education. A higher Traditional Beliefs score reflects strict and conservative beliefs about childrearing and education.

Training Support: No information available.

Adaptations/Special Instructions for Individuals with Disabilities: No information available.

Report Preparation Support: No information available.

References:

Administration for Children and Families. "Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start. Volume II: Final Technical Report Appendixes." Washington, DC: U.S. Department of Health and Human Services, 2002.

Schaefer, Earl S., and Marianna Edgerton. "Parent and Child Correlates of Parental Modernity." In *Parental Belief Systems: The Psychological Consequences for Children*, edited by I.E. Sigel. Hillsdale, NJ: Erlbaum, 1985.

Vogel, Cheri A., Kimberly Boller, Yange Xue, Randall Blair, Nikki Aikens, Andrew Burwick, Yevgeny Shrago, Barbara L. Carlton, Lara Kalb, Linda Mendenko, Judy Cannon, Sean Harrington, and Jillian Stein. "Learning as We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children." Report No. 2011-7. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2011.

PARENTING ALLIANCE MEASURE (PAM), 1999 (2011 Update)

Authors:

Richard R. Abidin and Timothy R. Konold

Publisher:

Psychological Assessment Resources, Inc.

(800) 331-8378

<http://www4.parinc.com>

Instrument:

<http://www4.parinc.com/Products/Product.aspx?ProductID=PAM>

Initial Material Cost:

PAM Introductory Kit (includes manual and 50 test forms for hand scoring): \$150

Representativeness of Norming Sample:

The norming sample was collected in 1997 and included 1,224 parents of children 1 to 19 years old (mean age 10.9 years) from 15 states. Parent median income ranged from \$40,001 to \$50,000. The sample resembled 1992 U.S. Census Bureau projections for 1997 with regard to ethnicity, parent marital status, and children's gender. Normative data were derived separately for mothers ($n = 879$) and fathers ($n = 345$) based on significant differences in PAM scores (fathers scored higher).

Languages:

English

Type of Assessment:

Parent report

Age Range and Administration Interval:

Parents of children 1 to 19 years old

Personnel, Training, Administration, and Scoring Requirements:

Little training is required to administer and score the PAM.

Authors recommend that assessors study the manual administration and scoring procedures and that someone with graduate-level training in psychology or a related field interpret the results.

The PAM takes 5 to 8 minutes to complete and fewer than 5 minutes to score.

Summary:

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (most .65 or higher)

Validity: 3 (most .5 or higher for concurrent)

Norming Sample Characteristics: 2 (older than 15 years)

Ease of Administration and Scoring: 2 (self-administered and scored by someone with basic clerical skills)

Description: The Parenting Alliance Measure (PAM) is a self-administered questionnaire that measures the perceived strength of the childrearing alliance between parents. It includes 20 items and may be administered to parents of children 1 to 19 years old. Although the PAM does not include subscales, the authors designed the PAM to reflect four dimensions of a positive parenting alliance as defined by Weissman and Cohen (1985) such that the parent (1) is invested

in the child, (2) values the other parent's involvement with the child, (3) respects the other parent's judgment, and (4) wants to communicate with the other parent.

Other Languages: None.

Uses of Information: Clinicians working with families use the PAM to discern whether parents cooperate with each other to meet the needs of their children. Researchers may use the PAM to

assess the relationship between parental behaviors and the experiences and behaviors of children.

Reliability:

1. Internal consistency: The authors found Cronbach's alpha coefficients of .96, .97, and .97 for fathers, mothers, and the total normative sample, respectively.
2. Test-retest reliability: coefficients for 27 fathers, 33 mothers, and all 60 parents were .63, .88, and .80, respectively, when the PAM was administered at a four- to six-week interval.
3. Inter-rater reliability: not applicable.

Validity:

1. Content validity: the PAM is based on an earlier measure developed by the same authors (Parenting Alliance Inventory), with 80 items developed by the authors and field experts. The authors field tested the 80-item measure with parents for readability and clarity, subjected items to expert review to rate whether they measured the intended construct, and reduced the measure to 30 items based on ratings. The authors further reduced the number of items to 20 after conducting factor analysis with 512 racially and socioeconomically diverse parents in central Virginia. Principal Component Analyses, Maximum Likelihood Analyses, and Principal Axis Analyses conducted on mothers and fathers in the normative sample resulted in one- and two-factor solutions. The authors adopted the one-factor solution

because factors were highly correlated and the second factor accounted for a small percentage of the score variance.

2. Concurrent validity: the PAM scores were compared to scores on the Stress Index for Parents and Adolescents (SIPA) and the Dyadic Adjustment Scale (DAS) measuring marital relationship quality as well as to the Family Adaptability and Cohesion Evaluation Scales III (FACES-III) Family Adaptability and Cohesion subscales. The authors included parents of adolescents in both the PAM norming sample (n = 713) and the SIPA norming sample (n = 64); the parents were primarily white, married, well educated, and reported higher incomes. The PAM and SIPA scores were correlated at -.55 for mothers and -.68 for fathers in the PAM norming sample and at -.72 for mothers and -.75 for fathers in the SIPA norming sample, indicating that higher parental stress was related to lower-quality parenting alliances. Among parents from the SIPA norming sample, the PAM and DAS scores correlated at .52 and .75 for mothers and fathers, respectively. The PAM and FACES-III Adaptability scores were correlated at .43 for fathers (correlation was non-significant for mothers), and the PAM and FACES-III Cohesion scores were correlated at .59 and .75 for mothers and fathers, respectively. The authors indicated that perceived parenting alliances and family cohesion and adaptability were more important to fathers' view of their family.

Authors compared the PAM scores to teacher ratings of children's self-esteem and social competence, as measured by the Behavioral Academic Self-Esteem (BASE) measure and California Preschool Social Competency Scale (CPSCS) (n = 160 parents and 78 teachers). For mothers and fathers, the PAM and BASE scores correlated at .20 and .29, respectively. The PAM and CPSCS were correlated at .32 for fathers (the relationship was non-significant for mothers, which, the authors suggested, could be attributable to mothers' more consistent actions toward children regardless of the perceived parenting alliance).

The authors reported statistically significant PAM score differences as measured by t-test comparisons and one-way analysis of variance between the normative sample and a clinical sample (n = 272 children 7 to 19 years old). Parents of children (1) without clinical diagnoses, (2) no history of receiving mental health services, and (3) no history of delinquent behavior had higher PAM scores than parents of children with clinical diagnoses or a history of mental health services or delinquency. Married parents had higher PAM scores regardless of whether their children were clinically diagnosed with a mental health disorder such as attention deficit hyperactivity disorder.

Bias Analysis: No information available.

Methods of Scoring: Parents rate each item on a five-point scale ranging from strongly agree to strongly disagree, and their responses are

recorded on a carbon-copy sheet. The PAM Total raw score is the sum of the 20 item scores. If items are missing responses (up to two missing responses are allowed), then the manual provides instructions for computing an average raw score. Percentiles and T scores for mothers, fathers, and in total are included in the manual.

Interpretability: The PAM Score Summary area on the scoring sheet indicates how percentile score ranges correspond to score interpretations: within normal limits, marginal, problematic, and dysfunctional. The manual includes case study examples for users interpreting the PAM; the case studies include family background, test results, and applications in family counseling settings.

Training Support: None; the authors indicate that the PAM does not require formal training to administer and score.

Adaptations/Special Instructions for Individuals with Disabilities: A grade 3 reading level is required for parents to self-administer the PAM. However, items may be read aloud to parents who have difficulty reading.

Report Preparation Support: The scoring sheet includes a PAM Score Summary area to record the total score, percentile, T score, and interpretation (within normal limits, marginal, problematic, dysfunctional).

References:

Abidin, Richard R., and Timothy R. Konold.
PAM Parenting Alliance Measure Professional

Manual. Lutz, FL: Psychological Assessment Resources, Inc., 1999.

Carlson, Cindy I. "Review of the Parenting Alliance Measure." In *The Fifteenth Mental Measurements Yearbook*, edited by Barbara S. Plake, James C. Impara, and Robert A. Spies. Lincoln, NE: The Buros Institute of Mental Measurements, 2003.

Clarke, Mary M. "Review of the Parenting Alliance Measure." In *The Fifteenth Mental Measurements Yearbook*, edited by Barbara S. Plake, James C. Impara, and Robert A. Spies. Lincoln, NE: The Buros Institute of Mental Measurements, 2003.

Weissman, S.H., and R.S. Cohen. "The Parenting Alliance and Adolescence." *Adolescent Psychiatry*, vol. 12, 1985, pp. 24-45.

PARENTING INTERACTIONS WITH CHILDREN: CHECKLIST OF OBSERVATIONS LINKED TO OUTCOMES (PICCOLO), 2009 (2011 Update)

Authors:

Lori Roggman, Gina Cook, Mark Innocenti, Vonda Jump Norman, and Katie Christiansen

Publisher:

Utah State University, Center for Persons with Disabilities; available by contacting Lori Roggman or Mark Innocenti
(435) 797-0091 or (435) 797-2006
<http://www.cpdusu.org/projects/piccolo/>

Initial Material Cost:

None; the measure and supporting materials are currently available free of charge from the authors.

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

1 through 3 years old

Personnel, Training, Administration, and Scoring Requirements:

Administered by a highly trained program staff member or researcher

Assessors receive group training from the authors or other experienced PICCOLO trainers. The authors recommend at least 1 day of training, which includes a 2-hour period of observing and coding videos to demonstrate inter-rater reliability with the authors.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (most are .65 or higher)

Validity: 2 (all under .4 for concurrent and predictive)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO) is an observational tool designed to measure positive parenting behaviors as parents interact with their toddlers 1 through 3 years old. The measure includes 29 items organized by four domains: (1) Affection (7 items), (2) Responsiveness (7 items), (3) Encouragement of Autonomy (7 items), and (4) Teaching (8 items). The authors recommend administering the PICCOLO with videotaped parent-child interactions, particularly for research use, although it may be administered “live” with parents and toddlers. The authors also recommend that parents and toddlers interact with stimulating materials (e.g., books, puzzles,

pretend-play toys) during the observation. The PICCOLO may be incorporated into observations of family routines or planned home visit activities. Furthermore, the Early Head Start Family and Child Experiences Survey (Vogel et al. 2001) applies the PICCOLO coding scheme to videotaped observations of the Two Bags Task, in which one bag contains a picture book and the other contains cooking- and food-related toys.

Other Languages: None.

Uses of Information: Coding the videos in the presence of a parent or program supervisor enables practitioners to highlight the behaviors that contributed to the ratings. Practitioners and program staff may use data from the PICCOLO to

provide feedback to families as part of parent education and continuous program improvement efforts. Others may use the data for research, evaluation, program monitoring, and development purposes.

Reliability:

1. Internal consistency reliability: Cronbach's alpha was reported for each domain: Affection was .78, Responsiveness was .75, Encouragement was .77, and Teaching was .80.
2. Test-retest reliability: Pearson's correlation coefficients for the domains ranged from .39 to .52 for a 10-month administration interval (14 to 24 months) and from .36 to .51 for a 12-month interval (24 to 36 months).
3. Inter-rater reliability: groups of at least three observers coded 2,300 videos, and the overall inter-rater reliability was .84 (coefficient type not specified). The inter-rater reliability analysis included data from an additional 2,200 videotaped observations coded by at least two coders. Across more than 4,500 videos, average percentage agreement was reported for each domain: Affection was 80 percent, Responsiveness was 76 percent, Encouragement was 83 percent, and Teaching was 69 percent.

Validity:

1. Content validity: the PICCOLO was developed by using videotaped observations of parents and children engaged in the Three Bags Task. The initial set of PICCOLO items was drawn from a review of child

development theory and constructs linked to child outcomes in Head Start, Early Head Start, low-income populations, observational measures of parent-child interactions, and other Head Start program materials. With the initial set of items identified, Head Start program staff and parent volunteers, along with untrained observers, reviewed the items to assess each item's clarity and usefulness. The authors based final item selection on reliability and validity results from coded, videotaped observations and qualitative feedback from the coders and Head Start partners. A confirmatory factor analysis on the final instrument indicated that item loadings were in the moderate to high range, with .43 to .86 for Affection, .55 to .78 for Responsiveness, .62 to .70 for Encouragement, and .58 to .69 for Teaching. The authors asked practitioners in two Early Head Start programs and in one home visiting program—all of whom were trained in the PICCOLO—to evaluate the degree of each item's importance based on a three-point scale ("somewhat important" to "very important"). The average importance rating for the final set of PICCOLO items was 2.6 compared to 2.3 for the items eliminated from the measure.

2. Concurrent validity: PICCOLO scores were correlated with several child outcomes.¹ Among 14-month-olds, Affection, Encouragement, and Teaching domain scores

¹ The PICCOLO technical report did not provide information on which versions of the measures were used (e.g., Peabody Picture Vocabulary Test, Fourth Edition or previous version).

were correlated with the cognitive development scale of the Scales of Infant Development Mental Development Index (MDI), and coefficients ranged from .12 to .14. Among 24-month-olds, all PICCOLO domain scores were correlated with the MDI, and coefficients ranged from .21 to .27. PICCOLO Affection, Encouragement, and Teaching domain scores were correlated with the emotion regulation scores on the Behavior Rating Scales (BRS) of the Bayley Scales of Infant Development-Second Edition, and coefficients ranged from .14 to .15. These three PICCOLO domains were also correlated with the aggression score on the Child Behavior Checklist (CBC) at -.12 for each. Scores on the PICCOLO Responsiveness, Encouragement, and Teaching domains were correlated with the Vocabulary Production scale on the Communication Development Inventories (CDIs), and coefficients ranged from .13 to .22. Among 36-month-olds, all PICCOLO domain scores were correlated with MDI, and coefficients ranged from .16 to .22. All PICCOLO domain scores at 36 months were correlated with Peabody Picture Vocabulary Test (PPVT) scores, and coefficients ranged from .13 to .18.

3. Predictive validity: PICCOLO scores were predictive of several child outcomes.¹ PICCOLO domain scores at 14 months were predictive of MDI scores at 24 months (correlation coefficients ranged from .19 to .27), CDIs scores at 24 months (.13 to .17), PPVT scores at 36 months (.16 to .22), PPVT

scores at pre-kindergarten (.20 to .26), emergent literacy scores on the Woodcock Johnson Letter Word Identification test (WJLW) at pre-kindergarten (.15 to .17), and problem-solving scores on the Woodcock Johnson Applied Problems test (WJAP) in pre-kindergarten (.13 to .19). PICCOLO Affection scores at 14 months predicted BRS scores at 24 months at .12. PICCOLO domain scores at 24 months were predictive of MDI scores at 36 months (.18 to .28), PPVT scores at 36 months (.20 to .25), PPVT scores at pre-kindergarten (.24 to .30), WJLW scores at pre-kindergarten (.14 to .18), and WJAP scores at pre-kindergarten (.16 to .21). PICCOLO domain scores at 36 months were predictive of PPVT scores at pre-kindergarten (.22 to .27), WJLW scores at pre-kindergarten (.16 to .22), and WJAP scores at pre-kindergarten (.13 to .16).

4. Construct validity: PICCOLO domains are moderately to highly correlated with one another among toddlers 14 to 36 months old (n=1020 to 1616). Intercorrelation coefficients between Teaching and other domains (Affection, Responsiveness, and Encouragement) were lower (.40 to .60) than coefficients between Affection, Responsiveness, and Encouragement (.65 to .73). This pattern held across age groups.

Bias Analysis: The authors tested inter-rater reliability on a subsample of more than 500 videos of black, white, and Hispanic toddlers as a function of the video coders' race/ethnicity. Each

video was first coded by an assessor of the same race/ethnicity as the toddler. Then, two or more coders of different race/ethnicities as both the toddler and the original coder observed and coded the same video. For example, a Hispanic toddler was first videotaped and coded by a Hispanic assessor and then coded by white and black assessors. Ratings were compared by using percentage agreement by domain (i.e., three comparisons per domain for each toddler race/ethnicity). Percentage agreement ranged from 60 percent (Encouragement, black toddler/coder compared to other coder ratings) to 82 percent (Affection, white toddler/coder compared to other coder ratings), with the remaining comparisons ranging from 64 to 79 percent.

Methods of Scoring: Each item is scored as 0 = absent (behavior not observed), 1 = barely (brief, minor, or emerging behavior observed), or 2 = clearly (definite, strong, or frequent behavior observed). Assessors add scores for each item to calculate a domain score. No overall score is calculated.

Interpretability: The authors provide a scoring grid that may be used to interpret domain scores by age. The grid summarizes scores indicative of high risk, moderate risk, and strength (e.g., low/no risk) for each domain such that risk corresponds to suboptimal toddler development. The scoring grid is helpful for identifying families' strengths and areas that need improvement.

Training Support: The authors are available to train researchers and practitioners in the

PICCOLO, and group training is available. The PICCOLO project web site indicates that DVD training materials are forthcoming.

Adaptations/Special Instructions for Individuals with Disabilities: No information available.

Report Preparation Support: No information available.

References:

Cook, Gina A., and Lori A. Roggman. "PICCOLO (Parenting Interactions with Children: Checklist of Observations Linked to Outcomes) Technical Report." Logan, UT: Utah State University, 2009.

Roggman, Lori A., Gina A. Cook, Mark S. Innocenti, Vonda K. Jump, and Katie Christiansen. "PICCOLO (Parenting Interactions with Children: Checklist of Observations Linked to Outcomes) User's Guide." Logan, UT: Utah State University, 2009.

Vogel, Cheri A., Kimberly Boller, Yange Xue, Randall Blair, Nikki Aikens, Andrew Burwick, Yevgeny Shrago, Barbara L. Carlton, Lara Kalb, Linda Mendenko, Judy Cannon, Sean Harrington, and Jillian Stein. "Administration for Children and Families. Learning as We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children. OPRE Report #2011-7." Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2011.

PARENTING STRESS INDEX, THIRD EDITION (PSI), 1995

Authors:

Richard R. Abidin

Publisher:

Psychological Assessment Resources, Inc.

(800) 331-8378

www.parinc.com

Initial Material Cost:

PSI Long Form Kit: \$131 (includes manual, 10 reusable item booklets, 25 hand-scorable answer sheet/profile forms)

PSI Short Form Kit: \$90 (includes manual, 25 hand-scorable questionnaire/profile forms)

Representativeness of Norming Sample:

The English version of the PSI was standardized with parents of children ranging from 1 month to 12 years (mean of 4.9). The non-random sample of parents included 2,633 mothers (ages ranging from 16 to 61, with a mean of 30.9) and 200 fathers (ages ranging from 18 to 65, with a mean of 32.1). The parents were recruited by clinic, school, or child care center staff and volunteered to participate in the norm sampling study. The Spanish version was normed on a sample of 223 Hispanic parents.

Languages:

English, Spanish, French

Type of Assessment:

Parent report

Age Range and Administration Interval:

For parents of children ages 1 month to 12 years

Personnel, Training, Administration, and Scoring

Requirements: The manual states that an individual without formal training in psychology or social work can administer and score the PSI, but the interpretation of PSI scores requires someone with training in these or other related disciplines. Parent needs to have at least a 5th grade education. The PSI long form takes approximately 20 to 30 minutes to complete.

Summary

Initial Material Cost: 2 (\$100 to \$200)

Reliability: 3 (internal consistency and test-retest .65 or higher for both the Long and Short Forms)

Validity: 3 (mostly .5 or higher for concurrent validity)

Norming Sample Characteristics: 2 (not nationally representative)

Ease of Administration and Scoring: 3 (no special administration requirements, scored by highly trained individual)

Description: The purpose of the 120-item PSI is to produce a diagnostic profile of perceived child and parent stress. The PSI was developed based on the theory that total parental stress is a function of child and parent characteristics, as well as situational variables. It contains 13 subscales within 4 major domains: total stress, child domain, parent domain, and life stress. The total stress domain, which measures the level of stress in the parent-child relationship, is comprised of the child and parent domains. The child domain

has six subscales that measure the child's distractibility/hyperactivity, adaptability, reinforcement of the parenting experience, demandingness, mood, and acceptability. The remaining seven subscales make up the parent domain and measures: competence, isolation, attachment, health, feeling of role restriction, depression, and spousal support. The life stress domain measure sources of stress beyond the parent's control. The PSI is also available in a

Short Form, which consists of a 36-item self-scoring questionnaire and profile.

Uses of Information: Primary uses are screening for early identification, assessment for individual diagnosis (including informing therapy and counseling), pre-post measurement for effectiveness of intervention, and research for studying the effects of stress on parent-child interactions.

Reliability: (1) Internal consistency (Cronbach's alpha) for the PSI (Long Form) subscales ranged from .70 to .83 in the Child Domain, .70 to .84 in the Parent Domain, and was greater than .90 for the two domains and the Total Stress scale. Similar internal consistency alphas for the PSI were also established in a cross-cultural population study (Hauenstein, et al., 1987). In the PSI Short Form (PSI/SF) subscales, internal consistency (Cronbach's alpha) was .85 in the Difficult Child, .80 in Parent-Child Dysfunctional Interaction, .87 in Parental Distress, and .91 in Total Stress. (2) Test-retest reliabilities (intervals between administrations of the PSI in these studies ranged from 3 weeks to 1 year) in the PSI Long Form ranged from .55 to .82 for the Child Domain, .69 to .91 for the Parent Domain, and .65 to .96 for the Total Stress score.

Validity: (1) Concurrent validity: the manual provides an abstract of studies that demonstrated concurrent validity by comparing PSI (Long Form) scores with those on other assessment instruments. Only a few of the abstracts provided validity statistics. The few that reported statistics found that the correlation between Total Stress

and the Bayley Scale was .42 at 3 months and .66 at 6 months. The correlation between child domain and negative behavior in hyperactive siblings relationships was .60, while its correlation with the 6 factors in the Family Impact Questionnaire ranged from .36 to .84. A study also reported correlations ranging from .65 to .77 between life stress and the lack of formal support among parents of children with disabilities.

Method of Scoring: The PSI contains a hand-scorable Answer Sheet on which basic demographic information and item responses are included. Most responses require the respondent to circle SA (strongly agree), A (agree), NS (not sure), D (disagree), or SD (strongly disagree) in response to the particular items. Addition and, if there are missing data, division skills are needed to obtain the raw scores. Using the profile form, which is part of the answering sheet, the scorer can obtain the percentile ranking for each subscale score. The respondent's score can also be graphed on the profile form. Detailed information on scoring is provided in the Professional Manual.

The PSI also offers a Software Portfolio, Windows software that allows you to administer either the 120-item PSI or the 36-item PSI Short Form on-screen or to enter item responses from the PSI or the PSI Short Form. The software automatically scores the item responses and generates a report. Reports can be edited on-screen. This updated software program contains modifications to the interpretive statements, empirically based cutoff scores, and reference lists of PSI research.

Interpretability: The manual states that interpretation of the PSI scores requires someone who has graduate training in clinical, counseling, or educational psychology or in social work or a related field. Interpretation guidelines are discussed in the manual, and it is suggested that the individual reviewing and interpreting the results first interpret the Total Stress score, and then look at the Child Domain and Parent Domain scores and their subdomains scores to pinpoint the sources of stress. Throughout the interpretation guidelines in the manual, there are references to research literature. The interpretation section also includes five case illustrations profiling different parental and situational characteristics.

Training Support: None mentioned in manual

Adaptations/Special Instructions for Individuals with Disabilities: While there are no explicit instructions for administering the PSI with parents of children with disabilities, the manual contains information on how the PSI may

work with this population. Sections entitled “Families with Special-Needs Children” and “Disabilities and Illnesses” cites various research studies related to use of the PSI in families with children having some disabilities. These studies are summarized, and cover various disabilities including: autism, deafness, congenital heart disease, asthma, cystic fibrosis, and so forth.

Report Preparation Support: The software generates a report. Two sample reports can be found at the publisher’s website: www.parinc.com.

References:

Abidin, Richard R. Parenting Stress Index, Third Edition. Odessa, FL: Psychological Assessment Resources, 1995.

Hauenstein, E., S. Scarr, and Richard Abidin. Detecting Children At-risk for Developmental Delay: Efficacy of the Parenting Stress Index in a Non-American Culture. Unpublished manuscript. Charlottesville: University of Virginia, 1987.

SUPPORT FUNCTIONS SCALE (SFS), 1985

Authors:

Carl J. Dunst and Carol M. Trivette

Publisher:

Brookline Books
(617) 558-8010 or
(800) 666-BOOK
www.brooklinebooks.com/

Initial Material Cost:

Book: Enabling and Empowering Families, \$25
\$10 per batch of 10 score sheets.

Representativeness of Norming Sample:

No norming sample. The research sample consisted of 121 parents of preschool mentally challenged, handicapped, and developmentally at-risk children.

Languages:

English

Type of Assessment:

Parent self-report

Age Range and Administration Interval:

Families of young children

Personnel, Training, Administration, and Scoring

Requirements: Parent (or caregiver) can complete the scale in 10 minutes. An early intervention practitioner can review the answers and interpret the scale, identifying places for concern, in under 10 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 2 (.65 or higher for internal, split-half, and test-retest reliability for individual items; < .65 for test-retest reliability for total scale score)

Validity: 2 (< .5 for criterion validity)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (self-administered)

Description: The self-report Support Functions Scale (SFS) measures the extent of parents' needs for different types of support. The scale is available in both an extended (20-item) and short (12-item) version. Both versions ask parents to rate their need for financial, emotional, instrumental, and informational support on a five-point scale ranging from never (1) to quite often (5).

Uses of Information: Providers can use the SFS results to guide follow-up discussions with parents to better understand their needs and to develop an intervention plan to address needs. The book provides a number of case studies to illustrate how the information can be used.

Reliability: (1) Internal consistency reliability (Cronbach's alpha): .87; split-half reliability (using the Spearman-Brown formula): .88; (2) Test-retest reliability (1-month interval): the average correlation among administrations for the individual items was .91; for the total scale score, the correlation among administrations was .62.

Validity: (1) Concurrent validity: The authors compared the outcomes on the SFS scale to the outcomes on scales measuring family well-being (McCubbin et al.), personal well-being (Trivette and Dunst, 1985), and time demand on respondent (Dunst and Trivette, 1985). The total scores (20-item scale) proved to be the best predictor. Both family (correlation = .27) and personal (correlation = .33) well-being were

significantly related to adequacy of support, whereas lack of support placed more time demands upon the respondent (correlation = -.20). Financial support was the only factor score significantly related to family well-being (correlation = .27), whereas emotion (correlation = .17), child-related (correlation = .21), and instrumental (correlation = .29) support were significantly related to personal well-being. None of the factor scores were related to the personal time demands measure.

Methods of Scoring: The parent or caregiver answers to what extent he or she feels a need for each type of assistance by marking Never, Once in a While, Sometimes, Often, and Quite Often.

Interpretability: If the respondent rates an item as Sometimes, Often, or Quite Often (have a need), this may be taken as an indication that further interviewing (assessment) is necessary to determine the exact type of help that is needed but lacking.

Training Support: None. However, a second book published by Brookline Books, entitled "Supporting & Strengthening Families: Methods Strategies and Practices" is a collection of papers updating the thinking and practices described in "Enabling and Empowering Families", and building and elaborating upon the model described in the earlier book.

Adaptations/Special Instructions for Individuals with Disabilities: None, this measure was developed based on work with families with children who have disabilities.

Report Preparation Support: The book provides a Family Support Plan form and a Profile of Family Needs and Support form for the agency to use. The needs and support form can be used to identify providers and record the resources they are expected to provide to help the family address an identified need. The family support form helps the agency to mobilize resources to address needs and to monitor the progress of the intervention.

References:

Dunst, Carl and Carol Trivette. Personal Time Commitment Scale: Reliability and Validity. Unpublished scale. Morganton, NC: Family, Infant and Preschool Program, 1985.

Dunst, Carl, Carol Trivette, and Angela Deal. Enabling and Empowering Families: Principles & Guidelines for Practice. Cambridge: Brookline Books, 1988.

Dunst, Carl, Carol Trivette, and Angela Deal. Supporting and Strengthening Families: Methods, Strategies, and Practices. Newton: Brookline Books, 1994.

McCubbin, H.I., J.K. Comeau, and J.A. Harkins. "Family Inventory of Resources for Management." In H.I. McCubbin and J.M. Patterson (eds.), Systematic Assessment of Family Stress, Resources, and Coping. St. Paul, MN: Family Stress and Coping Project, 1981, pp. 67-69.

Trivette, Carol and Carl Dunst. Personal Well-Being Index: Reliability and Validity. Unpublished scale. Morganton, NC: Family, Infant and Preschool Program at Western Carolina Center, 1985.

Program Implementation and Quality Instruments



Program Implementation and Quality Instruments^a

Instrument	Domains	Age Range	Assessment Type	Initial Material Cost	Reliability	Validity	Norming Sample	Ease of Administration & Scoring
Arnett Caregiver Interaction Scale	Child care quality	0-5.5 years	2	1	3	3	1	3
Classroom Assessment Scoring System (CLASS) Toddler and Pre-K	Child care quality	15-60 months	2	3	3	2	1	3
Early Childhood Environment Rating Scale-Revised	Child care quality	2.5-5 years	2	1	3	1	1	3
Early Head Start Evaluation—Parent Services Interviews	Program services	NA	3	1	1	1	1	1
Family Child Care Environment Rating Scale-Revised Edition (FCCERS-R)	Child care quality	0-12 years	2,3	1	3	1	1	3
Head Start Family Information System ^b	Program services, including home visit documentation	NA	3	1	1	1	1	3
Home Visit Rating Scales (HOVRS) and Home Visit Rating Scales-Adapted (HOVRS-A)	Program quality	0-2 years	2	1	3	1	1	3
Infant/Toddler Environment Rating Scale Revised Edition (ITERS-R)	Child care quality	0-2.5 years	2,3	1	3	3	1	3
National Association for the Education of Young Children Accreditation Criteria	Child care quality	NA	3	1	3	NA	NA	1
The Parent-Caregiver Relationship Scale (PCRS)	Program quality	2-24 months	3	1	3	2	1	2
Program Implementation Checklist and Rating Scales Developed for the National Early Head Start Research and Evaluation Project	Program quality	NA	3	1	1	NA	1	3
Program Review Instrument for Systems Monitoring (PRISM), 2002	Program quality	NA	2,3	1	1	1	1	1

KEY

Assessment Type

- 1 = Direct assessment
- 2 = Observation
- 3 = Parent/self report

Initial Material Cost

- 1 = Under \$100
- 2 = \$100 to \$200
- 3 = More than \$200

Reliability

- 1 = None described
- 2 = Under .65
- 3 = .65 or higher

Validity

- 1 = None described
- 2 = Under .5 for concurrent; under .4 for predictive
- 3 = .5 or higher for concurrent; .4 or higher for predictive

Norming sample

- 1 = None described
- 2 = Older than 15 years, not nationally representative or representative of EHS population
- 3 = Normed within past 15 years, nationally representative or representative of EHS population

Ease of administration and scoring

- 1 = Not described
- 2 = Self-administered or administered and scored by someone with basic clerical skills
- 3 = Administered and scored by a highly trained individual

^a The information included in this table was drawn from the manuals or other resources available from the authors and publishers of the instruments. Individual users may have different experiences.

^b The HSFIS is a management information system. It also includes measures of home-based service use and forms for tracking child and family outcomes.

ARNETT CAREGIVER INTERACTION SCALE, 1989¹

Authors:

Jeffery Arnett

Publisher:

None.

A copy of the scale can be found in Jaeger and Funk (2001)

Cost:

None

Representativeness of Norming Sample:

None described.

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

Caregivers of early childhood classes

Personnel, Training, Administration, and Scoring Requirements:

To be a certified Arnett Caregiver Interaction Scale observer requires achieving a .70 inter-rater reliability coefficient for two consecutive visits. (Jaeger and Funk). No recommended length of observation. Arnett observed caregivers in two 45-minute sessions, while Jaeger and Funk observed caregivers in a 2.5-hour session.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: Internal consistency and inter-rater reliability: 3 (.65 or higher)

Validity: Concurrent: 3 (mostly .5 or higher)

Norming Sample Characteristics: 1 (none described)

Administration and Scoring: 3 (administered and scored by a highly trained individual)

¹ The scale is also referred to as the Arnett Scale of Caregiver Behavior.

Description: The 26-item Caregiver Interaction Scale assesses the quality and content of the teacher's interactions with children. The scale was designed to provide information on various socialization practices that have been identified in research on parenting. The scale can be used without modification in both center and home-based settings. The items measure the emotional tone, discipline style, and responsiveness of the caregiver in the classroom. The items are usually organized into the following four sub-scales: (1) positive interaction (warm, enthusiastic, and developmentally appropriate behavior), (2) punitiveness (hostility, harshness, and use of threat), (3) detachment (uninvolvement and disinterest), and (4) permissiveness.

Uses of Information: The scale can be used to assess caregiver's interactions with children and their emotional tone and approach to engaging and disciplining children.

Reliability: (1) Internal consistency: Layzer et al. obtained Cronbach alphas of .91 for warmth/responsiveness (positive interaction) and .90 for harshness (punitiveness), while Resnick and Zill obtained alphas for the total scale of .98 for lead teachers and .93 for assistant teachers. Jaeger and Funk reported coefficients of .81 and higher for the sensitivity (positive interaction), punitiveness, and detachment subscales. (2) Inter-rater reliability: Jaeger and Funk reported inter-rater reliability coefficients ranging from .75 to .97 between a certified observer and trainees.

Validity: (1) Concurrent validity: Layzer et al. reported correlation coefficients of .43 to .67 between the Arnett and the Early Childhood Environment Rating Scale (ECERS), Assessment Profile for Early Childhood Programs, and the Description of Preschool Practices. The authors did not expect the coefficients to be large because the Arnett scale focused more narrowly on an aspect of teacher behavior not directly measured by the other three observation instruments. However, Phillipsen et al. reported a correlation of .76 between the Arnett and the ECERS.

Method of Scoring: The observer rates the extent to which the caregiver exhibits the behavior described in the item on a 4-point scale, ranging from not at all (1) to very much (4). Averages can be calculated for each subscale.

Interpretability: Depending on the program's needs, individual caregiver scores can be compared to the scores of other caregivers or the mean scores of a group of caregivers compared against the means of other groups of caregivers. Statistical tests have been frequently utilized to assess the differences between scores.

Training Support: None described.

Adaptations/Special Instructions for Individuals with Disabilities: None described.

Report Preparation Support: None described.

References:

Arnett, Jeffery. "Caregivers in Day-Care Centers: Does Training Matter?" *Journal of*

Applied Developmental Psychology. Vol. 10, 1989, pp. 541-552.

Jaeger, Elizabeth, and Suzanne Funk. *The Philadelphia Child Care Quality Study: An Examination of Quality in Selected Early Education and Care Settings*. Philadelphia: Saint Joseph's University, October 2001.

Layzer, Jean I., Barbara D. Goodson, and Marc Moss. *Observational Study of Early Childhood Programs, Final Report, Volume I: Life in Preschool*. Cambridge, MA: Abt Associates, Inc., 1993.

Phillipsen, Leslie, Debby Cryer, and Carollee Howes. "Classroom Process and Classroom Structure." In *Cost, Quality, and Child Outcomes in Child Care Centers*, edited by Suzanne W. Helburn. Denver: Department of Economics, Center for Research in Economics and Social Policy, University of Colorado at Denver, 1995, pp. 125-158.

Resnick, Gary, and Nicholas Zill. *Is Head Start Providing High-Quality Education Services? "Unpacking" Classroom Processes*. Albuquerque, NM: Biennial Meeting of the Society for Research in Child Development, April 15-18, 1999.

U.S. Department of Education. National Center for Education Statistics. *Measuring the Quality of Program Environments in Head Start and Other Early Childhood Programs: A Review and Recommendations for Future Research*, Working Paper No. 97-36, by John M. Love, Alicia Meckstroth, and Susan Sprachman. Jerry West, Project Officer. Washington, DC: 1997.

CLASSROOM ASSESSMENT SCORING SYSTEM (CLASS) TODDLER AND PRE-K, 2008 (2011 Update)

Authors:

Robert C. Pianta, Karen M. La Paro, and Bridget K. Hamre

Publisher:

Paul H. Brookes Publishing Co.
(800) 638-3775
<http://www.brookespublishing.com>

Instrument:

<http://www.brookespublishing.com/resource-center/screening-and-assessment/class/>

Initial Material Cost:

Toddler CLASS: Unpublished

CLASS Pre-K Manual: \$49.95; 10 booklets each with 6 observations sheets and 1 scoring sheet: \$28

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

Toddler CLASS is administered in classrooms to toddlers 15- to 36-months-old. CLASS Pre-K is administered in preschool classrooms.

Personnel, Training, Administration, and Scoring

Requirements: Administered by a highly trained program staff member. Extensive training is required to achieve and maintain acceptable levels of reliability for CLASS results.

CLASS Pre-K: Introductory training costs \$210 per person or \$2,000 for groups of 8 to 25 people. Observer training costs \$670 per person for two days. On-site training for 7 to 15 people costs approximately \$4,805 (material costs may be lower depending on the number of trainees). Observers recertify annually for \$35. Three-day train-the-trainer sessions cost \$2,100 per person or \$6,000 for groups of 6 to 10 people. The costs above exclude personal travel and lodging (regional training) and trainer travel and lodging and costs associated with the host facility (on-site training).

The CLASS takes at least 2 hours to administer and score.

Summary:

Initial Material Cost: 3 (> \$200)

Reliability: 3 (mostly .65 or higher)

Validity: 2 (mostly under .5 for concurrent)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual).

Description: The Classroom Assessment Scoring System (CLASS) is an observation tool for assessing child care classroom quality among classrooms with infants (Infant CLASS), toddlers 15 to 36 months old (Toddler CLASS), preschoolers (CLASS Pre-K), and students in kindergarten through grade 12 (CLASS K-3, CLASS Upper Elementary, and CLASS

Secondary).¹ Observations are conducted in person or by videotaping; results are tallied within scales on an Observation Sheet. Observers complete four observation cycles that

¹ The CLASS Upper Elementary and CLASS Secondary are in the pilot phase. The Infant CLASS and Toddler CLASS are under development. The Infant CLASS is not described because little information is available.

take 20 minutes per cycle; they also take notes and require 10 minutes to score results. Observed practices and behaviors are scored as minimally to highly characteristic of the classroom on a rating scale of low (1, 2), mid (3, 4, 5), and high (6, 7).

The Toddler CLASS and CLASS Pre-K focus on the quality of interactions between teachers and children along three scales (termed domains by the authors): Emotional Support, Classroom Organization, and Instructional Support. Each scale includes several dimensions. Emotional Support dimensions are Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Child Perspectives. Classroom Organization dimensions are Behavior Guidance (Toddler CLASS) and Behavior Management, Productivity, and Instructional Learning Formats (CLASS Pre-K). Instructional Support dimensions are Quality of Feedback and Language Modeling in both CLASS versions as well as Facilitation of Learning and Development (Toddler CLASS) and Concept Development (CLASS Pre-K).

Other Languages: None

Uses of Information: Using a common metric and vocabulary, the Toddler CLASS and CLASS Pre-K standardize the description of the quality of the classroom environment across the early childhood period. Researchers may use the two measures to study classroom quality by focusing on administrators in order to assess their accountability to stakeholders and to undertake program planning and evaluation and

by focusing on programs in order to provide teachers with feedback on improved instructional quality.

Reliability:²

1. Internal consistency reliability: The Emotional Support scale for the Toddler CLASS had a Cronbach's alpha coefficient of .88 (Thomason and La Paro 2009). Unpublished analyses from the nationally representative Baby FACES longitudinal study of 89 Early Head Start programs (Baby FACES study) (n = 220 classrooms with 2-year-olds) reported Cronbach's alpha coefficients of .79 (Classroom Organization), .84 (Emotional Support), and .95 (Instructional Support) (C. Vogel, Personal Communication, June 2011).
2. Test-retest reliability: the stability of the CLASS Pre-K scales and their dimensions among the National Center for Early Development and Learning Multi-State

² Reliability and validity information is drawn from four studies. The Toddler CLASS observations were used in a 2007 pilot study in a Southeastern state with 46 female teachers from 30 classrooms with 3 to 15 children and 1 to 6 teachers per classroom (Thomason and La Paro 2009). The CLASS Pre-K was used in (1) the National Center for Early Development and Learning Multi-State Prekindergarten Study (NCEDL MS Pre-K) of 240 randomly selected pre-kindergarten classrooms across six states, observed between 2001 and 2002; (2) the NCEDL State-Wide Early Education Programs (SWEET) Study of 500 randomly selected pre-kindergarten classrooms across five states, observed between 2003 and 2004; and (3) the MyTeachingPartner (MTP) Study of 164 classrooms from which teachers agreed to participate among randomly selected Virginia districts.

Prekindergarten Study (NCEDL MS Pre-K) classrooms are available within observations and across time periods. Scale correlations across the four cycles of the observation ranged from 0.86 (Instructional Support) to .91 (Emotional Support). Dimension coefficients ranged from .79 (Instructional Learning Formats) to .90 (Teacher Sensitivity). Over two consecutive days, scale coefficients ranged from .81 (Classroom Organization) to .86 (Instructional Support). Dimension coefficients ranged from .73 (Productivity) to .85 (Teacher Sensitivity). Between fall and spring, dimension coefficients ranged from .25 (Quality of Feedback) to .64 (Behavior Management).

3. Test-retest reliability: the stability of the CLASS Pre-K scales and their dimensions among the National Center for Early Development and Learning Multi-State Prekindergarten Study (NCEDL MS Pre-K) classrooms are available within observations and across time periods. Scale correlations across the four cycles of the observation ranged from 0.86 (Instructional Support) to .91 (Emotional Support). Dimension coefficients ranged from .79 (Instructional Learning Formats) to .90 (Teacher Sensitivity). Over two consecutive days, scale coefficients ranged from .81 (Classroom Organization) to .86 (Instructional Support). Dimension coefficients ranged from .73 (Productivity) to .85 (Teacher Sensitivity). Between fall and spring, dimension

coefficients ranged from .25 (Quality of Feedback) to .64 (Behavior Management).

Validity:²

1. Content validity: Development of the CLASS was based on literature reviews, focus groups, and pilot testing. Researchers also drew on the precursor to the CLASS, the Classroom Observation System (COS) for children in grades 1, 3, and 5. To inform the development of the Toddler CLASS, researchers reviewed existing measures (Infant/Toddler Environment Rating Scale, Caregiver Interaction Scale, and Observational Record of the Caregiving Environment), and infant and toddler experts reviewed Toddler CLASS itself. CLASS Pre-K development was supported by consultations with experts on classroom quality and teaching effectiveness and by pilot testing in the NCEDL MS Pre-K Study. Confirmatory factor analysis for the CLASS Pre-K investigated three factors representing Emotional Support, Classroom Organization, and Instructional Support. Analysis of the MTP Pre-K, NCEDL MS Pre-K, and the NCEDL State-Wide Early Education Program (SWEET) studies showed factor loadings at or above .70 except for two factor loading coefficients for Classroom Organization dimensions (.56 for Productivity from the MTP Pre-K Study and .66 for Instructional Learning Formats in the SWEET Study).

2. Confirmatory factor analysis for the CLASS Pre-K investigated three factors representing Emotional Support, Classroom Organization, and Instructional Support. Analysis of the MTP Pre-K, NCEDL MS Pre-K, and the NCEDL State-Wide Early Education Program (SWEEP) studies showed factor loadings at or above .70 except for two factor loading coefficients for Classroom Organization dimensions (.56 for Productivity from the MTP Pre-K Study and .66 for Instructional Learning Formats in the SWEEP Study).
3. Predictive Validity: combined results from the NCEDL MS Pre-K and SWEEP studies indicated that classroom quality is linked to student academic and social gains between fall and spring of the preschool year (Howes et al. 2008). The Instructional Support scale was the most robust predictor of growth. Classrooms in which the teacher promoted higher-order thinking and creativity and provided oral feedback (that is, higher Instructional Support scale scores) showed an increase in preschoolers' receptive and expressive language skills (effect sizes of .06 and .07, respectively) based on Hierarchical Linear Modeling HLM analyses controlling for maternal education, ethnicity, and gender. Mashburn and colleagues (2008) showed that instructional quality, when controlling for pretest scores, child and family characteristics, and state, was positively associated with children's receptive language, expressive language,

rhyming, applied problem solving, and letter naming. The quality of emotional interactions was positively related to the development of social competence and negatively related to the development of problem behaviors.

4. Construct validity: Toddler CLASS researchers examined aspects of toddler development and observed toddler child care classrooms. Correlations were significant between measure dimensions and established correlates of classroom quality, including classroom ratings, program star ratings, teacher education level, teacher membership in a professional organization, group size, and teacher-child ratio.

Bias Analysis: No information available.

Methods of Scoring: Each dimension receives a score on a seven-point scale (1 = minimally characteristic to 7 = highly characteristic) for each of the four cycles in the observation. Composite scores for each dimension are average scores over the four cycles. Dimension composite scores within a scale are then averaged to yield a scale score that indicates whether interactions fall in the low, middle, or high range.³ Given that scoring relies on raters to conduct observations of long segments of instruction, scoring is highly subjective, and authors recommend that raters refer to the manual before giving scores. The

³ For the Emotional Support scale, the Negative Climate dimension score must be reversed (subtract composite score from 8) before using it to calculate the scale score.

professional training program offered by the CLASS authors is needed to attain high levels of reliable scoring.

Interpretability: The manual and CLASS training sessions provide extensive instruction in scoring practices and interpretation. The manual also includes examples of scores in each range. Higher scores for the scales indicate higher-quality interactions between teachers and students.

Training Support: Several training options are available for the CLASS Pre-K. Trainees attend regional trainings (schedule is available at classobservation.com) or host on-site trainings. Introductory training covers CLASS structure and uses, and it is offered in two-, four-, and six-hour sessions. Observer training familiarizes researchers and teachers with data collection practices and provides practice in observing and coding data through pre-recorded sessions and a reliability evaluation. Trainees are then certified to observe and code classrooms with the CLASS. Trainees must achieve an average 80 percent agreement (within one point) across five reliability DVDs, with at least two scores per scale correct across the five DVDs. Additional support is available for participants who do not pass the reliability evaluation at the end of the training period. Three-day train-the-trainer sessions teach individuals how to train other potential CLASS users.

Subscriptions to the CLASS Video Library are available on classobservation.com. The library offers video clips illustrating high-quality

teacher-student interactions and teaching examples in pre-kindergarten classrooms.

Adaptations/Special Instructions for Individuals with Disabilities: Not applicable.

Report Preparation Support: No information available.

References:

Halle, Tamara, Jessica E. Vick Whittaker, and Rachel Anderson. “*Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition.*” Washington, DC: Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2010.

Howes, Carollee, Margaret Burchinal, Robert Pianta, Donna Bryant, Richard M. Clifford, and Oscar Barbarin. “Ready to Learn? Children's Pre-Academic Achievement in Pre-Kindergarten Programs.” *Early Childhood Research Quarterly*, vol. 23, no. 1, 2008, pp. 27-50.

Mashburn, Andrew, Robert Pianta, Bridget K. Hamre, Jason T. Downer, Oscar Barbarin, Donna Bryant, Margaret Burchinal, Richard M. Clifford, Diane Early, and Carollee Howes. “Measures of Classroom Quality in Prekindergarten and Children's Development of Academic, Language, and Social Skills.” *Child Development*, vol. 79, no. 3, 2008, pp. 732-749.

Pianta, Robert C., Karen M. Paro, and Bridget K. Hamre. "The CLASS Elementary Video Library." Available at [<http://store.teachstone.org/class-video-library/>].

Pianta, Robert C., Karen M. La Paro, and Bridget K. Hamre. *Classroom Assessment Scoring System: Pre-K Manual*. Baltimore: Brookes Publishing Co., 2008.

Pianta, Robert C., Karen M. La Paro, and Bridget K. Hamre. *Classroom Assessment Scoring System: Toddler Version*. Unpublished.

Thomason, A. C. and Karen M. La Paro. "Measuring the Quality of Teacher-Child Interactions in Toddler Child Care." *Early Education and Development*, vol. 20, no. 2, 2009, pp. 285-304.

EARLY CHILDHOOD ENVIRONMENT RATING SCALE - Revised Edition (ECERS-R), 1998

Authors:

Thelma Harms, Richard M. Clifford, and Debby Cryer
www.fpg.unc.edu

Publisher:

Teachers College Press
www.teacherscollegepress.com
1-800-575-6566

Initial Material Cost:

ECERS-R Assessment Scale: \$13;
Video training: \$59;
Workbook: \$4

Representativeness of Norming Sample:

Research reported in manual does not include this information

Languages:

English¹

Type of Assessment:

Observation, with some caregiver report

¹ Please contact Thelma Harms to obtain information about other official translations.

Age Range and Administration Interval:

For classrooms enrolling children 2 1/2 to 5 years of age.

Personnel, Training, Administration, and Scoring Requirements:

The individual administering the ECERS-R must read and practice the scale, and must also have knowledge of child development and educational implications. The authors recommend reviewers to have at least two practice observation sessions with an experienced ECERS-R trainer. Administration time ranges from 2 hours to 5 hours depending on the scoring option selected for administering the assessment.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (.65 or higher for internal consistency)

Validity: 1 (concurrent not available, predictive is promising)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The ECERS-R is designed to assess group programs for children of preschool age (2 1/2 to 5). It is a 43-item assessment tool rating scale organized into seven environmental subscales: Space and Furnishings, Personal Care Routines, Language-Reasoning, Activities, Interaction, Program Structure, and Parents and Staff. Each item has a number of quality indicators. The ECERS-R can be used in preschool, kindergarten, and child care classrooms. The original ECERS was revised to reflect changes in the early childhood field and to be more inclusive of children with disabilities and sensitive to cultural diversity.

Uses of Information: The assessment can be used by program directors for supervision and program improvement, by teaching staff for self-assessment, by agency staff for monitoring, and in teacher training programs.

Reliability: (1) Internal consistency (Cronbach's alpha) ranged from .71 to .88 at the subscale level; the total scale internal consistency was .92. (2) Inter-observer reliability was .92 (Pearson correlation) and .87 (Spearman correlation).

Validity: Concurrent validity information is not available in the manual. However, the original version of the ECERS has been demonstrated to

have good predictive validity (Peisner-Feinberg and Burchinal 1997), indicating that quality of center-based child care (measured using the ECERS-R) was related to preschool children's concurrent cognitive and socioemotional development (measured using the Peabody Picture Vocabulary Test-Revised and the Woodcock-Johnson Tests of Achievement-Revised).

Method of Scoring: The scoring sheet records the ratings for quality indicators, items, subscale and total scores, as well as any observer comments. The indicators, which have Yes/No/Not applicable response choices, are used to score the items, which have scores ranging from 1 (Inadequate) to 7 (Excellent). There are two ways to score the items. The manual provides detailed instructions on both scoring systems. The usual scoring system for each item is based on the number of the highest quality indicators with affirmative responses. Under the alternate scoring method, each indicator is individually scored (using the 1 to 7 range), which could extend the assessment time to a total of 4 to 5 hours. A Profile sheet is also provided to graphically display the scoring information, to compare areas of strengths and weaknesses, and to select items and subscales to target for improvement. The profiles for at least two observations can be plotted side by side to depict changes visually. Sample profile forms are available on the Frank Porter Graham Child Development Center website (www.fpg.unc.edu).

Interpretability: Full instructions for using the scale, plus notes clarifying selected scale items are included. However, the manual provides no information about interpreting the results of the observation.

Training Support: The assessment tool includes instructions for using the ECERS-R; a Video training package and workbook are also available. The website: www.fpg.unc.edu includes helpful information.

Adaptations/Special Instructions for Individuals with Disabilities: The revised version includes new items to assess program's services for children with disabilities. However, the revision needs field-testing and standardization.

Report Preparation Support: Not included in the manual.

References:

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

Harms, Thelma and Debby Cryer. Early Childhood Environment Rating Scale Video Observations, Revised Edition. New York: Teachers College Press, 1999.

Harms, Thelma and Debby Cryer. Early Childhood Environment Rating Scale Video Guide & Training Workbook, Revised Edition. New York: Teachers College Press, 1999.

Peisner-Feinberg, E., and M. Burchinal.
“Relations Between Preschool Children’s Child
Care Experiences and Concurrent Development:
The Cost, Quality and Outcomes Study.” Merrill-
Palmer Quarterly, vol. 43, no. 3, 1997, pp. 451-
477.

www.fpg.unc.edu

NATIONAL Early Head Start Research and Evaluation Project (EHSRE) PARENT SERVICES INTERVIEWS, 1996 - 2001

Authors:

John Love and other project staff, in collaboration with the Early Head Start Research Consortium

Publisher:

Mathematica Policy Research, Inc. Contact Publications, 609-275-2350, jallen@mathematica-mpr.com. The interviews can be found at www.mathematica-mpr.com or http://www.acf.hhs.gov/programs/opre/ehs/ehs_research/index.html.

Initial Material Cost:

None

Representativeness of Norming Sample:

None described

Languages:

English, Spanish

Type of Assessment:

Parent report

Age Range and Administration Interval:

For child-related questions, prenatal through 36 months.

For parent-related questions, all adults.

Personnel, Training, Administration, and Scoring

Requirements: The questions were designed to be administered to parents by someone with basic interviewing skills. Very little scoring is required.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 2 (administered by someone with basic interviewing skills)

Description: The Parent Services Interviews (PSIs) developed for the national Early Head Start Research and Evaluation Project (EHSRE) were designed to include instruments that assess potential program effects on service needs and use, as well as outcomes related to economic self-sufficiency. The interview questions were drawn from a variety of sources and include published instruments, questions drawn from other large national surveys, and questions that were developed specifically for this study. They cover topics including: family goals, perceived needs and resources, employment, education and job training, child care, home visits, transportation, housing, social support, health status, health care services, case management, other family support services, and public assistance receipt. The

interviews can be found on the web at the addresses listed above.

The results of the evaluation (through age 3) are included in two reports and their appendices and are available at <http://www.mathematica-mpr.com> and http://www.acf.hhs.gov/programs/opre/ehs/ehs_research/index.html.

Uses of Information: The Early Head Start PSIs can be used by programs to obtain service use information that can be compared to the national study findings. The summary information on services obtained from parents in the PSI questions may be especially useful to programs that do not collect detailed service use information in a management information system. For copyrighted instruments, programs must obtain

permission to use the assessments and must pay for their use.

Reliability: The technical appendices of the two reports include internal consistency reliability for summary scores. As a general rule, summary scores were not included in the report if their reliability was not above .65. Most measures in the PSI interviews were single questions and did not require computing summary scores.

Validity: Many of the questions were included in the evaluation because they had been used before in large studies and had demonstrated construct validity. Validity work based on the data collected was not reported in the two reports.

Method of Scoring: Most PSI measures were based on single questions and do not need scoring. Scoring procedures for any measure requiring scoring are summarized in the reports.

Interpretability: The information obtained from the PSI questions is easily interpretable.

Training Support: As part of the evaluation project, in-depth training manuals were developed; these can be obtained by requesting them from Jackie Allen at jallen@mathematica-mpr.com. In addition to the manuals, interviewers attended a central training session and had to meet rigorous standards before administering the study instruments. Mathematica is not providing any training support for the measures.

Adaptations/Special Instructions for Individuals with Disabilities: Contact

Mathematica for more information about how the protocols were adapted for use with individuals with disabilities.

Report Preparation Support: None described.

References:

ACYF. "Building Their Futures: How Early Head Start Programs Are Enhancing the Lives of Infants and Toddlers in Low-Income Families. Volume I: Technical Report." Washington, DC: Administration on Children, Youth and Families, 2001. www.mathematica-mpr.com or http://www.acf.hhs.gov/programs/opre/ehs/ehs_rsrch/index.html.

ACF. "Making a Difference in the Lives of Infants and Toddlers and Their Families: The Impacts of Early Head Start." Washington, DC: Administration for Children and Families, June 2002. www.mathematica-mpr.com or http://www.acf.hhs.gov/programs/opre/ehs/ehs_rsrch/index.html.

Sprachman, S., C. DeSaw, L. Mendenko, M. Salem, K. Boller, and B. Kolln. "Early Head Start National Evaluation Data Collection Training Manual for 6-Month Parents Services Interview, 14-Month Parent-Child Interview and Assessment, and Child Care Observations." Princeton, NJ: Mathematica Policy Research, Inc., December 1996.

www.mathematica-mpr.com

FAMILY CHILD CARE ENVIRONMENT RATING SCALE-REVISED EDITION (FCCERS-R), 2007 (2011 Update)

Authors:

Thelma Harms, Debby Cryer, Richard M. Clifford

Publisher:

Teachers College Press
(212) 678-3929
<http://www.tpress.com>

Instrument: <http://ers.fpg.unc.edu/node/111>

Initial Material Cost:

FCCERS-R (includes Rating Scale, Score sheet, and Instructions): \$20.95 (per unit)

Spanish version of scale also available: \$20.95 (per unit)

Representativeness of Norming Sample:

No norming sample

Languages:

English, Spanish

Type of Assessment: Observation, with some direct caregiver interview questions

Age Range and Administration Interval:

Birth to 12 years

Personnel, Training, Administration, and Scoring

Requirements: Administered by a trained observer. The authors recommend that observers attend a training session (with at least two classroom practice

observations) led by an experienced FCCERS-R trainer, followed by an inter-rater reliability comparison. Additional field practice observations may be necessary.

A Video Observations DVD (\$64) along with a Video Guide and Training Workbook (\$4) are available for observers. Authors at the University of North Carolina, Chapel Hill provide a three-day training course on how to use the instrument and conduct assessments for \$1,025.

The FCCERS-R typically takes 3.5 hours (3-hour observation, 20 to 30 minutes of questions). If all indicators are rated beyond the quality level score assigned to an item, then the observation would take 3 to 4 hours and questions would take 45 minutes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (mostly .65 or higher)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The Family Child Care Environment Rating Scale-Revised Edition (FCCERS-R) is used to assess the quality of family child care programs serving children birth through 12 years. Trained observers such as researchers, licensing personnel, evaluators, and program staff may administer the scale by making in-person observations and administering a short provider interview. The FCCERS-R contains 38 items, each of which consists of a set of numbered

statements (indicators) about the child care environment, including the space, interaction, activities, schedules, and provisions for parents and providers. The 460 indicators are distributed among seven subscales: Space and Furnishings, Personal Care Routines, Listening and Talking, Activities, Interaction, Program Structure, and Parents and Provider. Each item is rated from 1 to 7 with quality descriptors associated with the following levels: (1) inadequate (does not even

meet custodial care needs); (3) minimal (meets custodial needs and, to some extent, basic developmental needs); (5) good (meets developmental needs); and (7) excellent (high-quality personalized care).

The FCCERS-R was revised from the previous Family Daycare Rating Scale (FDCRS 1989), with several substantive and formatting changes. The FCCERS-R includes an additional subscale, Program Structure. Other changes include more culturally sensitive language, sample questions for difficult-to-observe indicators, and the replacement of several items with new ones that are more inclusive and address the needs of the full age range (instead of separate subitems for infants and older children). The earlier version also contained a separate supplementary subscale targeted to children with disabilities. In the revised version, the authors incorporated these indicators into the full scale and dropped the supplementary subscale.

Other Languages: A Spanish translation of the FCCERS-R, *Escala De Calificación Del Ambiente Ciudadano Infantil En Familia-Edición Revisada*, is available for purchase on the publisher's web site and is listed as approved by the authors. Detail on the norming sample, reliability, validity, and English language equivalence is unavailable.

Uses of Information: The authors indicate that the quality ratings from the FCCERS-R may be used in a wide variety of applications such as research, training, and technical assistance. Family

child care providers may also use the FCCERS-R to conduct a self-assessment of program quality.

Reliability: The authors conducted field tests across a sample of 45 family child care homes selected from six counties in central North Carolina, in a range of urban and rural settings. To ensure diverse representation, the sample was stratified by the star-quality ratings of the child care homes¹ and the ages of children served.

1. Internal consistency reliability: the Cronbach's alpha for the full FCCERS-R scale was .9. The authors also present alphas by subscales: .71 (Space and Furnishings), .46 (Personal Care Routines), .83 (Listening and Talking), .88 (Activities), .84 (Interaction), .62 (Program Structure), and .39 (Parents and Provider). The authors recommend caution when interpreting the reliability of two subscales, Personal Care Routines and Parents and Provider; the alphas are lower, and items within the subscales may be measuring different concepts.
2. Test-retest reliability: the two administrations had an 80.8 percentage agreement (.73 correlation coefficient, type not stated) across a subset of 20 child care homes. The mean interval between tests was 14 weeks.
3. Inter-rater reliability: the authors reported 88 percent agreement across all items, with a weighted kappa of .71. The percent agreement

¹ The state has devised a rating system to guide families on the quality of each home; ratings range from one star (meets only basic licensing requirements) to five stars (meets increasingly difficult requirements).

for the subscales ranged from 80 to 91 percent while the weighted kappas for the subscales ranged from .62 to .77.

Validity:

1. Content validity: the authors indicate that the FCCERS-R is based on four main sources: (1) research evidence from findings of the impact of child care environments on children's health and development; (2) content comparison between the FDCRS and other assessment instruments for similar age groups and settings; (3) feedback solicited from FDCRS users in a questionnaire; and (4) the authors' own experiences in using the scale.

The authors also relied heavily on the experiences of and findings from focus groups conducted during revisions of the Early Childhood Environment Rating Scale and Infant/Toddler Environment Rating Scales, both of which are related instruments that focus on center-based settings, especially items related to diversity and inclusivity.

Methods of Scoring: The assessor uses the score sheet to rate each item. Each indicator within an item is scored as Yes, No, or Not Applicable. Each item has a possible score of 1 to 7: (1) inadequate (does not even meet custodial care needs); (3) minimal (meets custodial needs and, to some extent, basic developmental needs); (5) good (meets developmental needs); and (7) excellent (high-quality personalized care). The Training Guide and Workbook provides instructions on how to assign the ratings to each

item. Users may calculate average subscale and total scale scores by summing the scores for each item and dividing by the number of items scored. The scoring sheet provides space for notes.

Interpretability: The scale includes a Profile, which allows for a graphic representation of all items and subscales. Assessors may use the Profile to compare areas of strength and weakness within a program and to target items or subscales for improvement.

Training Support: The authors recommend that observers attend a supervised training session (with at least two classroom practice observations lasting two to three hours each) led by an experienced FCCERS-R trainer, followed by an inter-rater reliability comparison. Additional field practice observations may be necessary. An interactive DVD/VHS of observations and an Instructor's Guide and companion Training Guide and Workbook are used during training. These materials are available for purchase separately through the Publisher's web site.

Adaptations/Special Instructions for Individuals with Disabilities: The manual instructs trained observers to ask the provider whether any children with disabilities are present in the facility before they begin their observations, but it does not require identification of the children. The Scale includes several indicators on Provisions for Children with Disabilities as part of the 38 items, with associated clarifications and questions for the provider.

Report Preparation Support: No information available.

References:

Harms, Thelma, and Richard M. Clifford. *Family Day Care Rating Scale*. New York: Teachers College Press, 1989.

Harms, Thelma, and Debby Cryer. *Video Guide and Training Workbook, Family Child Care Environment Rating Scale-Revised Edition*. New York: Teachers College Press, 2007.

Harms, Thelma, and Debby Cryer. *Video Observations for the Family Child Care Environment Rating Scale-Revised Edition*. New York: Teachers College Press, 2007.

Harms, Thelma, Debby Cryer, and Richard M. Clifford. *Family Child Care Environment Rating Scale-Revised Edition*. New York: Teachers College Press, 2007.

Harms, Thelma, Debby Cryer, Richard M. Clifford, and Ruy Burgos-Lovece. *Escala De Calificación Del Ambiente Ciudadano Infantil En Familia-Edición Revisada*. New York: Teachers College Press, 2010.

FAMILY DAY CARE RATING SCALE (FDCRS), 1989

Authors:

Thelma Harms and Richard M. Clifford
www.fpg.unc.edu

Publisher:

Teachers College Press
1-800-575-6566
www.teacherscollegepress.com

Initial Material Cost:

FDCRS Assessment: \$13
Video Observations: \$59
Video Guide and Training Workbook: \$4
30 Scoring Sheets: \$9

Representativeness of Norming Sample:

None described

Languages:

English¹

Type of Assessment:

Observation and self-assessment by program staff

Age Range and Administration Interval:

From 0 - 5 years

Personnel, Training, Administration, and Scoring Requirements:

Possible for supervisor, researcher, or trained day care provider to administer the FDCRS. Prior to administration, approximately 2 hours of reviewing the scale, 1 to 2 hours of video training, and two practice classroom observations (2 hours each) are recommended. FDCRS observation takes approximately 2 hours.

Summary:

Initial Material Cost: 1 (< \$100)
Reliability: 3 (.65 or higher)
Validity: 1 (concurrent not available, predictive is promising)
Norming Sample Characteristics: 1 (none described)
Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

¹ The FDCRS has also been translated into French. Those interested may contact Thelma Harms at the address above.

Description: The FDCRS is an adaptation of the Early Childhood Environment Rating Scale - ECERS (see ECERS-R review in this section), but focuses on the quality of family day care settings rather than center-based settings. The scale can be used by a supervisor, researcher, or trainer during an observation, or by a care provider as a self-assessment. The FDCRS contains 32 items organized in 6 subscales: Space and Furnishings for Care and Learning, Basic Care, Language and Reasoning, Learning Activities, Social Development, and Adult Needs. Each item is rated from 1 to 7 with quality descriptors associated with levels: (1) inadequate (does not even meet

custodial care needs); (3) minimal (meets custodial needs, and to some extent, basic developmental needs); (5) good (meets developmental needs); and (7) excellent (high-quality personalized care).

Uses of Information: The FDCRS was designed to be comprehensive, yet easy to use as part of supervision and monitoring by agency staff, self-evaluation by care providers, and also in research and program evaluation.

Reliability: (1) Internal consistency reliability (Cronbach's alpha) for the subscales: Space and Furnishings for Care and Learning (.86), Basic

Care (.90), Language and Reasoning (.90), Learning Activities (.93), Social Development (.83), and Adult Needs (.70); (2) Inter-rater reliability was .90 for individual items in studies reported by Howes and Stewart (1987) and Howes (1987). Reliability has been demonstrated for diverse groups in subsequent research.

Validity: (1) Concurrent validity: information not available in the manual. However, studies showed that scores on the earlier versions of the FDCRS were highly correlated (.80) with home visitors' ratings of family day care settings. The FDCRS is also related to caregiver education and child outcomes.

Method of Scoring: The packet includes a Score Sheet, to record scores for individual items. The Profile sheet permits a graphic representation of the scoring information. It can be used to compare areas of strengths and weaknesses, and to select items and subscales to target for improvement. The profiles for at least two observations can be plotted side by side to depict changes in a family child care home over time or differences between settings visually.

Interpretability: Full instructions for using the scale, plus notes clarifying selected scale items are included. However, the manual provides no information about interpreting the results of the observation.

Training Support: Video Observations and a Video Guide and Training Workbook are available for an additional cost. This multimedia package demonstrates how to use the Family Day

Care Rating Scale. Each training package contains an interactive videotape and an Instructor's Guide, which explains how to present the various training activities and provides answers and explanations for any questions that may arise. A 16-page Video Guide and Training Workbook contains training activities. The website: www.fpg.unc.edu also includes helpful information.

Adaptations/Special Instructions for Individuals with Disabilities: FDCRS includes Supplementary Items for Exceptional Children to be used when the facility enrolls a special needs child.

Report Preparation Support: Not included in the manual.

References:

Harms, Thelma, and Richard M. Clifford. Family Day Care Rating Scale. New York: Teachers College Press, 1989.

Harms, Thelma, and Richard M. Clifford. Family Day Care Rating Scale—Video Guide and Training Workbook. New York: Teachers College Press, 1993.

Harms, Thelma, and Richard M. Clifford. Family Day Care Rating Scale—Video Observations. New York: Teachers College Press, 1993.

Howes, C. Inter-Observer Reliability for the Harms and Clifford Family Day Care Rating Scale. Personal Communication, 1987.

Howes, C. and P. Stewart. "Child's Play with Adults, Toys, and Peers: An Examination of Family and Child Care Influences." *Developmental Psychology*, vol. 23, 1987, pp. 423-430.

www.fpg.unc.edu

HEAD START FAMILY INFORMATION SYSTEM (HSFIS) 4.3, 2001

Authors:

Head Start Bureau
www.acf.hhs.gov/programs/ohs/index.html

Publisher:

Cleverex Systems, Inc.
301-738-1122
www.cleverex.com
info@cleverex.com

Initial Material Cost:

HSFIS is available free of charge to Head Start programs.

Representativeness of Norming Sample:

None described.

Languages:

English, Spanish

Type of Assessment:

Management information system, including measures

of family needs, center-based services, and home-based services, as well as features for tracking child and family outcomes.

Age Range and Administration Interval:

For child-related information, prenatal to school age. For parent- and service-related questions, all adults.

Personnel, Training, Administration, and Scoring Requirements:

Varies. HSFIS is a management information system designed for use in tracking services and outcomes.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 3 (administered by a trained individual)

Description: The Head Start Family Information System (HSFIS) is an automated case management record-keeping system. It is designed to collect, organize, maintain, and report on information at both the child and family level to assist agencies in managing and improving services. The next generation HSFIS, a new web-based system called PROMIS (Program Resources and Outcomes Management Information System), has been developed and is being piloted in early 2003. It will include a module that can be used to track the progress and accomplishments of children in efforts to analyze and use data on child outcomes in program self-assessment and continuous improvement. It will be based on the framework set forth in ACF-HS-IM-00-18, "Using Child Outcomes in Program Self-Assessment."

Uses of Information: Information collected in HSFIS can help programs determine eligibility for enrollment, identify family service needs, provide program-level demographic statistics, track delivery of services to children and families, and track child and family outcomes.

Reliability: Not applicable.

Validity: Not applicable.

Method of Scoring: Not applicable.

Interpretability: Not applicable.

Training Support: Cleverex Systems, Inc. (<http://support.cleverex.com>, or 1-800-473-4780) provides training and technical assistance in the use of HSFIS and PROMIS. Cleverex Systems provides instruction at its own computer

laboratory in Rockville, Maryland and at local and regional training events. Training courses are offered at three levels—beginner, intermediate and advanced. Also, a 1-800 Help Desk provides technical assistance to users. Each Head Start Regional Office and Quality Improvement Center has a designated HSFIS liaison who can assist programs in obtaining automation funding and technical support.

Adaptations/Special Instructions for Individuals with Disabilities: None necessary.

Report Preparation Support: The HSFIS software includes the capability of producing more than 200 pre-defined reports. An Ad-Hoc Report Builder has been developed in the system

so that users can develop their own reports without additional training. In addition, advanced users can use Crystal Report software to develop more-sophisticated reports using HSFIS data.

References:

Angulo, Paulo. HSFIS Update.
<http://www1.cleverex.com/ShowArticle.asp?id=HSFIS1>. February 2000.

Gaidurgis, Andrew. Introducing PROMIS – Next Generation HSFIS.
<http://www1.cleverex.com/ShowArticle.asp?id=PROMIS>. May 2001.

HOME VISIT RATING SCALES (HOVRS), 2008 AND HOME VISIT RATING SCALES-ADAPTED (HOVRS-A), 2009 (2011 Update)

Authors:

HOVRS: Lori Roggman, Gina Cook, Vonda K. Jump Norman, Katie Christiansen, Lisa Boyce, and Mark Innocenti

HOVRS-A: Lori Roggman, Gina Cook, Vonda K. Jump Norman, Katie Christiansen, Lisa Boyce, Mark Innocenti, Nikki Aikens, Kimberly Boller, Diane Paulsell, and Kristin Hallgren

Publisher:

Brookes Publishing
Phone (800) 638-3775
<http://www.brookespublishing.com>

Instrument:

<http://products.brookespublishing.com/Developmental-Parenting-P566.aspx>

Initial Material Cost:

Developmental Parenting: A Guide for Early Childhood Practitioners (the appendix includes the measure):
\$26.95

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Observation

Age Range and Administration Interval:

0 to 24 months old

Personnel, Training, Administration, and Scoring Requirements:¹

Training for experienced observers lasts 3.5 days. It requires reliability certification by watching videos of home visits and matching their ratings to the gold standard on each scale three times.

Time to administer averages 1.5 hours during the home visit. Scoring time varies.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (all .65 or higher)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

¹ Training requirements are based on training sessions that were conducted by Mathematica Policy Research, Inc., in February 2009 for the Early Head Start Family and Child Experiences Survey (Vogel et al. 2011).

Description: The Home Visit Rating Scales (HOVRS) is an observation tool that evaluates the quality of home visits by assessing home visitor responsiveness, non-intrusiveness, support of parent-child interaction, and parent and child engagement during the visit. It may be used in homes with infants (0 to 12 months old) and toddlers (12 to 24 months old). Observers watch the home visitor in person or in video recordings and complete ratings for seven items. At the end of the observation, the observer records an

approximate percentage of time spent on each of the following types of activities: child-focused activities, parent-/family-focused activities, parent-child-focused activities, staff-family relationship-building activities, and crisis management activities. The HOVRS has two subscales: Home Visit Process Quality and Home Visit Effectiveness Quality. Home Visit Process Quality (four items) assesses the home visitor's responsiveness to family, relationship with family, facilitation of parent-child interaction, and non-

intrusiveness. Home Visit Effectiveness Quality (3 items) assesses parent-child interaction, parent engagement, and child engagement.

The HOVRS was adapted in several ways for the Early Head Start Child and Family Experiences Study (Baby FACES) and was subsequently named the Home Visit Rating Scales Adapted (HOVRS-A). The HOVRS-A allows for observers without home-visiting or clinical experience with families to use the measure. It includes fewer scoring categories to facilitate inter-rater reliability and adapts two items. The Home Visitor Relationship with Family item was adapted to capture the home visitor's engagement and relationship with the family and the family's relationship with the home visitor. The last item, Child Engagement during Home Visit, was adapted for two versions: one for visits with infants up to 12 months old and one for visits with toddlers 12 to 24 months old (Vogel et al. 2011).

Other Languages: None.

Uses of Information: The HOVRS describes and evaluates strategies used in home-visiting interventions, such as Partnering with Families for Early Learning (PEEL) in Washington as part of the Early Learning Initiative (Hallgren et al. 2010). It is also used as a measure to describe home-visiting quality in the Early Head Start Family and Child Experiences Survey (Vogel et al. 2011).

Reliability:

1. Internal consistency: for the HOVRS, Roggman and colleagues (Roggman et al. 2006) reported internal consistency

reliabilities for Visit Process Quality and Home Visit Effectiveness Quality subscales at .65, and the overall quality score had a Cronbach's alpha of .78. The internal consistency of the HOVRS-A as reported by Vogel and colleagues (2011) was .84 overall and .69 and .89 for the Home Visit Effectiveness Quality and Home Visitor Strategies subscales, respectively.

The HOVRS-A was used to observe children 1 week to 9 months old (average 3 months old) in 35 homes (Hallgren et al. 2010). Visits included a home visitor and the child's primary caregiver. Forty-three percent of the home visits were conducted in Spanish; the rest were conducted in English. The Cronbach's alphas for the total score, for the Home Visitor Strategies (i.e., Home Visit Process Quality), and for Participant Engagement (i.e., Home Visit Effectiveness Quality) were .87, .76, and .93, respectively.

2. Test-retest reliability: no information available.
3. Inter-rater reliability: no information available.

Validity:

1. Concurrent validity: the authors' unpublished findings from 59 observations in two home-based Early Head Start sites show that the HOVRS Home Visit Process Quality score predicts scores on the Parenting Interactions with Children: Checklist of Observations Linked to Outcomes (PICCOLO) and Home

Observation for Measurement of the Environment (HOME) Inventory (Roggman 2008).

2. Construct validity: the HOVRS-A overall and subscale scores correlated with features of the home visitors and the program (Vogel et al. 2011). Authors reported results of Chi-square tests and F-tests for categorical and continuous variable significance testing, respectively. Higher HOVRS-A scores were related to whether the home visitor has a state-awarded credential, higher risk of depression and less job satisfaction, and fewer unfilled staff positions. More time spent during the visit on parent-child activities and the involvement of fewer children in the observation are related to higher HOVRS-A scores.

Bias Analysis: No information available.

Methods of Scoring: HOVRS items are rated on a seven-point scale, with the anchor points at 1 (inadequate), 3 (adequate), 5 (good), and 7 (excellent). Observers review whether most items are checked as 1, 3, 5, or 7 to determine an overall rating between 1 and 7. Ratings for the Home Visit Process Quality scales may be summed to derive an index of process quality, and the Home Visit Effectiveness Quality scales may likewise be summed to create an index of effectiveness quality.

The HOVRS-A is rated on a five-point scale, with three anchor points at 1 (inadequate), 3 (adequate), and 5 (good). One item in the Home

Visitor-Family Relationship scale (Item 2: Home Visitor-Family Relationship) may be rated “yes” (1), “no” (0), or NA if family members other than the parent and child are not present during the home visit. Each anchor point is associated with indicators described in the manual. Observers use indicator instructions to determine ratings. For example, for an observer to assign a rating of 3, none of the indicators in anchor point 1 (inadequate) may be checked, and all indicators under 3 (adequate) must be checked. To assign a rating of 4, none of the indicators in 1 (inadequate) may be checked, all indicators under 3 (adequate) must be observed, and at least half of the indicators under 5 (good) must be observed. An overall quality rating for the home visit is based on the observer’s judgment and ranges from 1 (poor) to 5 (excellent).

Interpretability: No information available.

Training Support: HOVRS authors do not offer training. HOVRS-A authors provided training to accommodate research conducted for the Early Head Start Family and Child Experiences Survey (Vogel et al. 2011).

Adaptations/Special Instructions for Individuals with Disabilities: No information available.

Report Preparation Support: No information available.

References:

Hallgren, Kristin, Kimberly Boller, and Dianne Paulsell. “Better Beginnings: Partnering with Families for Early Learning Home Visit

Observations.” Report No. PP10-61. Princeton, NJ: Mathematica Policy Research, Inc., April 2010.

Hallgren, Kristen, and Jillian Stein. “Baby FACES Home Visit Rating Scales-Adapted (HOVRS-A) Manual (unpublished).” Princeton, NJ: Mathematica Policy Research, Inc., December 2009.

Roggman, L.A., K. Christiansen, G.A. Cook, V.K. Jump, L.K. Boyce, and C.A. Peterson. “Home Visits: Measuring How They Work.” Paper presented at the Early Intervention Research Institute Mini-Conference. Logan, UT, 2006.

Roggman, Lori. “Home Visit Observation Rating Scales (HVORS) Measurement Report” (unpublished report). 2008.

Roggman, Lori, Lisa Boyce, and Mark Innocenti. *Developmental Parenting: A Guide for*

Early Childhood Practitioners. Baltimore: Brookes Publishing, 2008.

Roggman, Lori A., Gina A. Cook, Vonda Jump, Lisa K. Boyce, and Mark S. Innocenti. “Home Visit Rating Scales (HOVRS)” (unpublished manuscript). Logan, UT: Utah State University, 2006.

Vogel, Cheri A., Kimberly Boller, Yange Xue, Randall Blair, Nikki Aikens, Andrew Burwick, Yevgeny Shrago, Barbara L. Carlton, Lara Kalb, Linda Mendenko, Judy Cannon, Sean Harrington, and Jillian Stein. “Learning as We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children.” Report No. 2011-7. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2011.

INFANT/TODDLER ENVIRONMENT RATING SCALE REVISED EDITION (ITERS-R), 2006 (2011 Update)

Authors:

Thelma Harms, Debby Cryer, and Richard M. Clifford

Publisher:

Teachers College Press

(800) 575-6566

<http://www.teacherscollegepress.com>

Instrument:

<http://store.tpress.com/0807746401.shtml>

Initial Material Cost:

ITERS-R Expanded Score Sheet and Profile for photocopying: \$20.95

Representativeness of Norming Sample:

No norming sample

Languages:

English, German, Japanese, and Spanish

Type of Assessment: Observation, with some direct caregiver interview questions

Age Range and Administration Interval:

For classrooms enrolling children from birth to 2.5 years old

Personnel, Training, Administration, and Scoring

Requirements: Administered by a trained observer.

The authors recommend that observers attend a

training session (with one or more practice observations) led by an experienced ITERS-R trainer. Researchers should contact the authors regarding training to evaluate inter-rater reliability. In addition, observers attending training should have knowledge of child development and educational implications (Frank Porter Graham Child Development Institute 2005).

ITERS-R training (excluding travel) at the University of North Carolina (UNC) ranges from \$1,025 to \$2,000 depending on the focus of the training. The Video Observations DVD is \$64, and the Video Guide and Training Workbook is \$4.

Takes 2 to 5 hours depending on scoring option. The average administration time is 3.5 hours, including the staff interview.

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (most are .65 or higher)

Validity: 3 (concurrent, applies to the original ITERS, which was given in percentage, not correlation.)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The ITERS-R is a classroom assessment tool designed to measure the quality of group programs for infants and toddlers (birth to 30 months) by collecting data through classroom observation and a staff interview. The assessment is a 39-item rating scale organized into seven environmental subscales: (1) Space and Furnishings (5 items), (2) Personal Care Routines (6 items), (3) Listening and Talking (3 items), (4) Activities (10 items), (5) Interaction (4 items), (6) Program Structure (4 items), and (7) Parents and

Staff (7 items). The items in the first six subscales are referred to as child-related, and the items in the last subscale are referred to as parent-/staff-related. Each item has several quality indicators, accounting for a total 467 Yes/No indicators. Reviewers caution against use of the assessment because it does not describe the validity of the measure in detail and instead relies on the validity research conducted on the 1990 version of the ITERS (Carey 2007; Kush 2007).

The original ITERS (Harms et al. 1990) and the revised versions exhibit several differences. The ITERS-R (Harms et al. 2003) updated the scoring for indicators to reflect observed strengths and weaknesses within items; removed negative indicators from all levels except for level 1 (Inadequate); lengthened the Notes for Clarification to enhance clarification; included culturally sensitive items and examples; added new items to some subscales (Listening and Talking, Activities, Program Structure, and Parents and Staff); combined items in the Space and Furnishings subscales in instances of apparent overlap; dropped two items from the Personal Care Routines subscale; and made more gradual the scaling of some items in the Personal Care Routines subscale. In addition, items appear on separate pages followed by the Notes for Clarification, and sample interview questions are included for difficult-to-observe indicators. The ITERS-R (Harms et al. 2006) features a new spiral binding, additional Notes for Clarification, and an Expanded Score Sheet, which includes notes and tables to assist with scoring. The items and indicators remain the same as in the 2003 version of the ITERS-R.

Other Languages: Despite the availability of official versions of the ITERS-R in other languages (German, Japanese, and Spanish), no information is available in the manual or the authors' web site regarding the development of the scales in those languages or investigations of comparability of scores with the English version.

Uses of Information: The assessment may be used by program staff as a self-assessment tool and by outside observers for program monitoring, evaluation, development, and research.

Reliability: Reliability studies apply to the ITERS-R (Harms et al. 2003). The items and indicators for the 2006 version of the ITERS-R are the same as for the 2003 version.

1. Internal consistency reliability: the total scale internal consistency was .93, and the internal consistency for the child-related items (items 1 through 32) was .92. Subscale internal consistency reliability ranged from .47 (Space and Furnishings) to .80 (Interaction), with four of the seven subscales at or above .70.
2. Test-retest reliability: no information available.
3. Inter-rater reliability: a two-phase pilot study conducted in 2001 and 2002 estimated the reliability of the 2003 version of the ITERS-R. The second study phase from which the reliability analysis was conducted included six trained observers who conducted 45 pairs of observations, each lasting about 3.5 hours (including the teacher report). Authors calculated interclass correlations, percentage agreement, and weighted Kappa statistics for inter-rater reliability. The interclass correlation was .92 for ratings based on the full scale as well as for the child-related items. Interclass correlations for ratings by subscale scores ranged from .67 (Personal Care Routines) to .92 (Parents and Staff). Authors

also calculated percentage agreement within 1 point for paired observations. There was agreement within 1 point 85 percent of the time across the full scale and 83 percent of the time across the 32 child-related items. Item agreement within 1 point ranged from 64 percent (Item 4: Room arrangement) to 98 percent (Item 38: Evaluation of staff). The weighted Kappa statistic for the full scale was .58 and .55 for the child-related scale. Two of the weighted Kappa statistics were below .40 (.14 for Item 9: Diapering/toileting and .20 for Item 11: Safety practices). Item 34 (Provisions for personal needs of staff) had the highest weighted Kappa statistics at .92. All Yes/No indicators achieved agreement of 91.7, and child-related indicators achieved agreement of 90.3. Item 11 (Safety practices) was the only item with indicator agreement below 80 (79.1 percent); and Item 35 (Staff professional needs) had the highest indicator agreement at 97.4.

Validity:

1. Content validity: to aid in classifying and assessing quality, the revision of the ITERS was based on research evidence from several relevant fields (e.g., health and education), best practices from professionals, and practical constraints of real life in child care settings. The revision process used four sources: (1) research on development in the early years and results associated with the impact of child care environments on children's health and development; (2) a

comparison of the content of the original ITERS with similar assessments and with documents describing program quality; (3) feedback via web site questionnaires and focus groups of professionals familiar with the ITERS; and (4) use of the ITERS over a two-year period by the co-authors and over 25 trained assessors for the North Carolina Rated License Project. Research and development provided information on the range of scores for certain items in addition to item-level difficulty and validity. The content comparison identified items to be added or eliminated. Revision of the original scale was based on results from the first phase of the reliability pilot study (see Inter-rater reliability); results from the second phase of the study resulted in the improvement of items with weighted Kappa statistics below .50 in order to improve reliability. The printed version of the scale specifies these changes.

A 2009 study of 223 observed Early Head Start classrooms included a principal components factor analysis (Vogel et al. 2011). The one-factor analysis findings were similar to those of Bisceglia et al. (2009) in that one factor showed a global aspect of quality and high internal consistency ($\alpha = .88$). Vogel et al. (2011) also found that a four-factor solution, including (1) Language/Interaction, (2) Activities, (3) Routines, and (4) Space/Furnishings, had alpha coefficients higher than .65 for each, which were more robust than established ITERS-R subscale coefficients.

2. Concurrent validity: the authors state that concurrent validity was established with the original version of the ITERS, in which a test for concurrent validity found 83 percent agreement between the categorizations of the quality of infant and toddler programs in 12 classrooms using the ITERS and expert evaluations. Given the similarity of the ITERS-R to the ITERS, studies on the ITERS-R have focused on the degree to which trained observers may continue to use the scale reliably (Harms et al. 2003).

ITERS-R subscales were correlated with measures of child development outcomes by using data from 367 classrooms of 1-year-old children in the Early Head Start Family and Child Experiences Survey (Vogel et al. 2001). Significant negative correlations demonstrated that higher program quality was associated with lower child development scores. The ITERS-R Activities subscale and Communicative Development Inventory (CDI) Spanish Comprehension scores correlated at $-.33$, and the ITERS-R Listening and Talking subscale score correlated with the CDI English Comprehension score at $-.14$. Authors suggested that negative correlations could be related to the greater likelihood that high-quality teachers and classrooms serve children at lower levels of development.

3. Predictive validity: the authors state that predictive validity was established with the original version of the ITERS. Given the similarity of the ITERS-R to the ITERS,

studies on the ITERS-R have focused on the degree to which trained observers may continue to use the scale reliably (Harms et al. 2003).

4. Construct validity: in an analysis of classrooms ($n = 211$ to 223) with 1-year-old children, ITERS-R scores were correlated with classroom characteristics associated with classroom quality (Vogel et al. 2001). Total ITERS-R scores were positively correlated with teachers' likelihood of returning to the classroom the following year ($.20$) and overall program implementation ($.20$). However, the authors were uncertain why total scores were correlated with teacher turnover ($.15$) and director/manager/coordinator turnover ($.17$) and negatively correlated with implementation of the family development cornerstone (deemed essential for program quality by the Advisory Committee on Services for Families with Infants and Toddlers) ($-.22$) and observed adult-child ratios ($-.23$).

Methods of Scoring: The Expanded Score Sheet is used to record the ratings for quality indicators, items, subscale scores, and total scores as well as any observer comments. The indicators, which have Yes/No/Not Applicable (NA) response choices, are used to score the items from 1 (Inadequate) to 7 (Excellent). Indicators fall under columns at the scale anchors 1, 3, 5, and 7. Items may be scored two ways as described in detail in the manual. Under the standard scoring option for each item, if any of the indicators in the

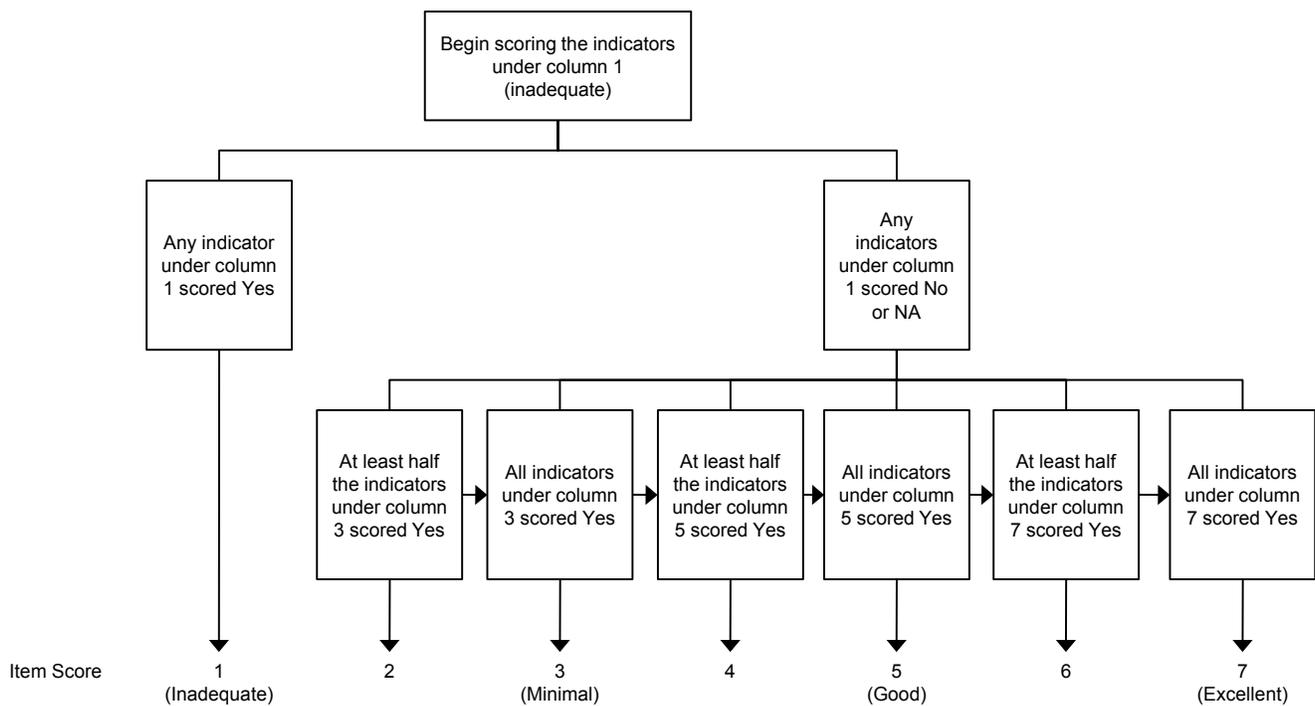
Inadequate column (or rating of 1) applies, the item is scored a 1. Higher item scores are determined by the number of indicators scored with a Yes response under each of the anchors, 3, 5, and 7 (Exhibit 1).

Under the alternative scoring method, each indicator is individually scored under each of the four anchors, potentially extending the assessment

time to a total of four to five hours. This scoring method is often used when the observation focuses on providing detailed feedback to programs or teachers.

Using either scoring method, subscale scores are calculated as the average rating across items for that subscale. The total score is calculated as the average item rating across all items.

Exhibit 1 Item Scoring Based on Indicators for the ITERS-R



Interpretability: Resources available for interpretation of scores include the Profile and the authors' web site. Observers must be thoroughly familiar with the ITERS-R. In addition, they are advised to be trained on the measure and demonstrate knowledge of child development and educational implications (Frank Porter Graham Child Development Institute 2005) before interpreting the results. Scoring information may be graphically displayed on the Profile to compare areas of strengths and weaknesses and to display items and subscales in need of improvement. The Profiles for two observations may be plotted side by side for visual depiction. A sample Profile appears in the manual along with blank Profile and Expanded Score Sheets for photocopying. In addition, the authors maintain an extensive web page (listed under Training Support) that answers questions about interpretability and use of the scale, and they have published a manual that goes beyond the information available in the instrument document.

Training Support: Individuals administering the ITERS-R should be highly trained. Training tools for the ITERS-R include the administration instructions in the manual, training aids from the publisher's web site, the *All About the ITERS-R* handbook, and in-person trainings. The Video Observations for ITERS-R, Instructor's Guide, Video Guide, and Training Workbook are available for purchase on the publisher's web site. The Video Observations for the ITERS-R DVD/VHS and the Instructor's

Guide demonstrate how to present training activities and answer frequently asked questions about the ITERS-R. In-person trainings are available during various times of the year. The authors' web site (<http://www.fpg.unc.edu/~ecers/>) provides information on in-person training and links to additional Expanded Score Sheets, Profiles, and other useful information. Observers participating in training sessions should demonstrate knowledge of child development and educational implications (Frank Porter Graham Child Development Institute 2005). In addition, researchers should contact the authors about training that includes evaluating inter-rater reliability with the authors. It is expected that the authors will train individuals or groups using the ITERS-R and will establish inter-rater reliability with them. Training of trainers and group leader training are available as is direct training of assessors. For those using the measures on a regular basis, the authors recommend re-establishing inter-rater reliability with them annually.

Adaptations/Special Instructions for Individuals with Disabilities: Not applicable.

Report Preparation Support: No information available.

References:

Bisceglia, R., M. Perlman, D. Schaack, and J. Jenkins. "Examining the Psychometric Properties of the Infant-Toddler Environment Rating Scale-Revised Edition in a High-Stakes

Context.” *Early Childhood Research*, vol. 24, 2009, pp. 121-132.

Carey, Karen. “Review of ITERS-R.” In *The Seventeenth Mental Measurements Yearbook*, edited by Kurt F. Geisinger, Robert A. Spies, Janet F. Carlson, and Barbara S. Plake. Lincoln, NE: The Buros Institute of Mental Measurements, 2007.

Frank Porter Graham Child Development Institute. “Environment Rating Scales.” Available at [<http://ers.fpg.unc.edu/>]. 2005.

Halle, Tamara, and Jessica Vick. “Quality in Early Childhood Care and Education Settings: A Compendium of Measures.” Submitted to the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Washington, DC: Child Trends, November 2007.

Harms, Thelma, and Debby Cryer. *Infant/Toddler Environment Rating Scale: Video Guide & Training Workbook*. New York: Teachers College Press, 2003.

Harms, Thelma, and Debby Cryer. *Video Observations for the Infant/Toddler Environment Rating Scale, Revised Edition*. New York: Teachers College Press, 2006.

Harms, Thelma, Debby Cryer, and Richard M. Clifford. *Infant/Toddler Environment Rating*

Scale, Revised Edition. New York: Teachers College Press, 2006.

Harms, Thelma, Debby Cryer, and Richard M. Clifford. *Infant/Toddler Environment Rating Scale, Revised Edition*. New York: Teachers College Press, 2003.

Harms, Thelma, Debby Cryer, and Richard M. Clifford. *Infant/Toddler Environment Rating Scale*. New York: Teachers College Press, 1990.

Harms, Thelma, Debby Cryer, and Cathy Riley. *All About the ITERS-R*. Lewisville, NC: Kaplan Early Learning Company, 2004.

Kush, Joseph C. “Review of ITERS-R.” In *The Seventeenth Mental Measurements Yearbook*, edited by Kurt F. Geisinger, Robert A. Spies, Janet F. Carlson, and Barbara S. Plake. Lincoln, NE: The Buros Institute of Mental Measurements, 2007.

Vogel, Cheri A., Kimberly Boller, Yange Xue, Randall Blair, Nikki Aikens, Andrew Burwick, Yevgeny Shrago, Barbara Lepidus Carlton, Lara Kalb, Linda Mendenko, Judy Cannon, Sean Harrington, and Jillian Stein. *Learning As We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children*. OPRE Report #2011-7. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2011.

NATIONAL ASSOCIATION FOR THE EDUCATION OF YOUNG CHILDREN (NAEYC) ACCREDITATION CRITERIA, 1998

Authors:

National Association for the Education of Young Children (NAEYC)

Publisher:

NAEYC
(202) 232-8777 or
(800) 424-2460
www.naeyc.org

Initial Material Cost:

Book: \$15
Accreditation fee varies by number of children: 60 or fewer children: \$575
61 to 120 children: \$900
121 to 240 children: \$1,100
241 to 360 children: \$1,250
\$200 for each additional 120 children

Representativeness of Norming Sample:

Not applicable

Languages:

English

Type of Assessment:

Program staff self-report, reviewed by NAEYC validators if accreditation is pursued

Age Range and Administration Interval:

Programs for young children of all ages. Accreditation good for 3 years.

Personnel, Training, Administration, and Scoring Requirements:

Not applicable

Summary:

Initial Material Cost: 1 (< \$100)
Reliability: 3 (95 percent agreement among reviewers)
Validity: Not applicable
Norming Sample Characteristics: Not applicable
Ease of Administration and Scoring: 1 (not described)

Description: The NAEYC accreditation system is designed to assist early childhood program personnel in making real and lasting improvements in quality of care and education, and to recognize programs that are in substantial compliance with the criteria for high-quality programs. NAEYC evaluates programs on the basis of their:

- Interaction among teachers and children
- Teaching
- Structure and processes used to facilitate the relationships among teachers and families
- Staff qualification and professional development opportunities
- Administration attention to the needs and desires of children, families, and staff

- Level of staffing
- Indoor and outdoor physical environment that facilitate learning
- Health and safety standards and practices
- Nutrition and food services
- Evaluation of program effectiveness.

NAEYC accreditation is a three-step process. The first step involves a self-evaluation, corrective actions, and completing the program description form that describes the program's compliance with the NAEYC criteria. The second step involves an on-site visit by NAEYC-trained validators to validate the accuracy of the information on the program description form. The third step involves the accreditation decision-making process by a three-person commission.

To be eligible for accreditation, the program must be licensed by the appropriate state/local agency or, if exempt from licensing, demonstrate compliance; serve at least 10 children younger than age 8; have at least two adults present at all times; and have been in operation for at least one year.

Uses of Information: Accredited programs may advertise their compliance with NAEYC standards. NAEYC provides accredited programs with a copy of the Commission Decision Report and their original validated program description, promotional materials, and a certificate indicating the expiration date of the accreditation. NAEYC provides programs with “deferred accreditation,” along with the specific reason for the deferral and recommendations for improvement. Programs can use the information to make improvements and request an additional on-site validation.

Reliability: NAEYC reports a 95 percent or higher agreement among commissioners reviewing the same program.

Validity: None described.

Method of Scoring: The three-person accreditation committee decides whether to grant accreditation or to defer accreditation until improvements can be made or additional information is obtained. The decision is based on the commissioners’ professional judgment and not on a point system. The commissioners review the validated program decision, taking into

account the context in which a program is operating and the overall impact of varying degrees of compliance for each component. To achieve accreditation, 100 percent compliance is not required; however, the program needs to be in substantial compliance with the accreditation criteria.

Interpretability: None described.

Training Support: Consultation by telephone and the Accreditation Criteria & Procedures of the National Academy of Early Childhood Programs and the Guide to Accreditation.

Adaptations/Special Instructions for Individuals with Disabilities: Not applicable.

Report Preparation Support: Not applicable.

References:

Accreditation Criteria and Procedures of the National Association for the Education of Young Children, 1998 Edition. Washington, DC: NAEYC.

Guide to Accreditation by the National Association for the Education of Young Children, 1998 Edition. Washington, DC: NAEYC.

Bredenkamp, Sue and Barbara Willer (Eds.). NAEYC Accreditation: A Decade of Learning and Years Ahead. Washington, DC: NAEYC, 1996.

Ethiel, Nancy (Ed.). Reflections on NAEYC Accreditation: Lessons Learned and Goals for the Future. Washington, DC: NAEYC, 1996.

PARENT CAREGIVER RELATIONSHIP SCALE (PCRS), 1996 (2011 Update)

Authors:

James Elicker, Illene C. Noppe, and Lloyd D. Noppe

Publisher:

James Elicker
(765) 494-2938
elickerj@purdue.edu

Initial Material Cost:

The PCRS includes 2 instruments: 1 Caregiver Scale and 1 Parent Scale. Instruments are available upon request from the authors at no charge.

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Parent report or caregiver report

Age Range and Administration Interval:

Parents and caregivers of children 2 to 24 months old

Personnel, Training, Administration, and Scoring Requirements:

Administered by parents and caregivers. The PCRS takes 10 to 20 minutes to administer.

No training required

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 3 (mostly .65 or higher)

Validity: 2 (all under .5 for concurrent)

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 2 (self-administered)

Description: The Parent Caregiver Relationship Scale (PCRS) was designed to measure the quality of the relationship between parents and caregivers in infant and toddler child care settings from the perspective of parents and caregivers. The PCRS is targeted to parents and caregivers of children 2 to 24 months old. It consists of two parallel questionnaires, each with 35 statements, about the parent-caregiver relationship; both a parent and the caregiver of the parent's child complete the questionnaires. Respondents rate their level of agreement or disagreement with each statement on a five-point scale. The two versions of the questionnaire, one with statements about the relationship tailored to parents and one with similar statements targeted to caregivers, consist of the following subscales: (1) Trust/Confidence, (2) Collaboration, and (3)

Affiliation (i.e., liking or friendship, in the Parent Scale) or Caring (in the Caregiver Scale).

Other Languages: None.

Uses of Information: The PCRS was designed to help researchers better understand the parent-caregiver relationship and to elucidate the perceptions of parents and caregivers about their relationships with each other.

Reliability: The authors conducted a field study in four Midwestern cities, drawing a sample from family child care (51 parents and 52 caregivers, yielding 42 respondent pairs) and center-based child care settings (73 parents and 41 caregivers, yielding 73 respondent pairs).

1. Internal consistency reliability: Cronbach's alphas were .93 for the Parent Scale and .94

for the Caregiver Scale. Alphas for the subscales ranged from .75 (Affiliation) to .91 (Confidence) for the Parent Scale and from .84 (Caring) to .92 (Confidence) for the Caregiver Scale.

2. Test-retest reliability: the authors calculated Pearson's correlations for samples in both family and center-based settings, with administration intervals between two and four weeks. In the family child care setting, coefficients were .80 for the Parent Scale and .84 for the Caregiver Scale. In the center-based child care setting, coefficients were .69 for the Parent Scale and .71 for the Caregiver Scale. Authors also reported that the Pearson's correlations for the subscales ranged from .59 to .78 (respondents not specified).
3. Inter-rater reliability: not applicable.

Validity:

1. Content validity: The authors conducted Principal Component Analysis (PCA) analyses on the full sample of 124 parents and 125 caregivers from the field study in four Midwestern cities. The analyses produced two similar factors for the parent and caregiver scales. The authors interpreted the factors as confidence and collaboration.
2. Concurrent validity: the authors correlated the total PCRS scores from Parent and Caregiver scales with several variables related to child care in respondent samples from center-based and family child care settings. In center-based settings, the authors found

positive correlations between the Caregiver Scale total scores with several variables: .27 (number of months in child care setting), .28 (caregiver work satisfaction), .33 (child care group size), .39 (caregiver rating of parent involvement in the center), and .45 (caregiver ratings of mothers' parenting skills). Parent Scale total scores were correlated at .27 with current hours per week in child care and .36 with parents' overall satisfaction with the child care arrangement.

The authors also reported statistically significant correlations between the variables described above and PCRS subscales. In the center-based subsample among parent respondents, the number of hours per week in care was most highly correlated with the Affiliation subscale (.30), and parents' satisfaction with child care was most highly correlated with the Collaboration subscale (.37). In the full sample of parent respondents, total time in all child care settings was negatively correlated with the parent Collaboration subscale (-.22).

For caregiver respondents, several variables were significantly correlated with the Confidence subscale, all in the center-based sample: child care group size (.33), caregivers' general perception of parents (.38), caregivers' rating of mothers (.48), caregivers' work satisfaction (.27), and number of hours per week in care (.25). Total time spent in child care settings was also significantly correlated with the Collaboration subscale.

Bias Analysis: The authors used t-tests to compare family child care and center-based child care subsample means for the PCRS total and subscale scores. The parent rating on the Caring subscale by the family caregiver sample was significantly higher than the center-based caregiver sample. Moreover, tested variables associated with child care, such as number of months in a child care setting and child care group size, were non-significant in the family child care subsample. Authors suggest that the lack of findings among the family child care subsample may be attributable to differences in the family child care environment (such as scheduling flexibility, amount of parent-caregiver interaction, and duration in care).

Methods of Scoring: Respondents (parents or caregivers) indicate their degree of agreement or disagreement with each questionnaire statement on a five-point scale: 1 (strongly disagree) to 5 (strongly agree). Item responses may be summed within subscales to obtain subscale scores, and scores across all subscales may be summed for the total PCRS scores.

Interpretability: The authors indicated that scores derived from the PCRS are intended to reflect the mutual perceptions of parents and

caregivers about their relationships in terms of providing care for infants. Authors intend to develop a “dyadic measure” that allows results to be combined for interpretation.

Training Support: No training required.

Adaptations/Special Instructions for Individuals with Disabilities: No information available.

Report Preparation Support: No information available.

References:

Elicker, James, Illene C. Noppe, and Lloyd D. Noppe. *The Parent-Caregiver Relationship Scale: Caregiver Version*. West Lafayette, IN: Purdue University, 1996.

Elicker, James, Illene C. Noppe, and Lloyd D. Noppe. *The Parent-Caregiver Relationship Scale: Parent Version*. West Lafayette, IN: Purdue University, 1996.

Elicker, James, Illene C. Noppe, Lloyd D. Noppe, and Cheryl A. Fortner-Wood. “The Parent-Caregiver Relationship Scale: Rounding Out the Relationship System in Infant Child Care.” *Early Education and Development*, vol. 8, no. 1, January 1997, pp. 83-100.

PROGRAM IMPLEMENTATION CHECKLIST AND RATING SCALES DEVELOPED FOR THE NATIONAL EARLY HEAD START RESEARCH AND EVALUATION PROJECT, 1997 (2011 Update)

Authors:

Ellen Kisker, Diane Paulsell, John Love, and Helen Raikes

Publisher:

Mathematica Policy Research, Inc.
Contact Publications, 609-275-2350,
jallen@mathematica-mpr.com, or visit the website,
www.mathematica-mpr.com

Initial Material Cost:

None

Representativeness of Norming Sample:

No norming sample

Languages:

English

Type of Assessment:

Program staff self-report

Age Range and Administration Interval:

Not applicable

Personnel, Training, Administration, and Scoring Requirements:

Not applicable

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (no norming sample)

Validity: Content validity established in relation to Head Start Program Performance Standards

Norming Sample Characteristics: 1 (no norming sample)

Ease of Administration and Scoring: 3 (administered and scored by a highly trained individual)

Description: The program implementation checklist and rating scales developed for the national Early Head Start Research and Evaluation project were designed to guide the collection and organization of information related to 25 key elements of the Head Start Program Performance Standards and Early Head Start program guidelines. The checklist can be used to guide the collection of information related to the key program elements, and the rating scales can be used to assess how fully the program has implemented each key element, how fully key program areas are being implemented, and how fully the program is being implemented overall. The checklist contains 39 general criteria with references to the performance standards, as well as specific indicators for each general criterion. Completing the checklist requires collecting

information from staff, parents, and program records. There are five ratings scales, one each for early childhood development and health services, family partnerships, staff development, community partnerships, and management systems. Each rating scale has multiple dimensions with ratings from 1 to 5. A rating of 4 indicates full implementation and a rating of 5 indicates enhanced implementation. The nationally representative Baby FACES longitudinal study of 89 Early Head Start programs (Baby FACES study) adapted the rating scales in 2009 (C. Vogel, personal communication, June 2011). Items were modified, and the administration was structured as a self-administered questionnaire to program directors.

Other Languages: None.

Uses of Information: The checklist and rating scales are intended to help Early Head Start programs serving pregnant women and families with infants and toddlers identify areas that need improvements in order to continue to provide high-quality comprehensive services that meet Head Start performance standards and other regulations. They may be useful for organizing information in preparation for the Office of Head Start monitoring visits.

Reliability: Reliability has not been established in the usual sense. However, an informal assessment conducted by the Office of Head Start monitoring staff concluded that the assessments of "full" or "enhanced" implementation on these rating scales were consistent with results of in-depth monitoring conducted by the Office.

Validity: Content validity was established by reviewing the specific criteria for determining the rating on each dimension with representatives of the Office of Head Start and the Early Head Start National Resource Center. A form of predictive validity was assessed in the national Early Head Start Research and Evaluation project, which found that programs that were rated as fully implemented achieved a stronger pattern of impacts on children and families (ACF 2003).

On the adapted rating scales used in the Baby FACES study in 2009, program directors' self-ratings were not always consistent with other information provided by programs in surveys and were not clearly associated with other indicators of program quality (C. Vogel, personal

communication, June 2011). As a result, researchers collected information in 2010 in the form of survey items and assigned response ranges and cut points in consultation with the Office of Head Start. Administration results are not available at this time.

Bias Analysis: None.

Methods of Scoring: The Early Head Start evaluation team developed overall ratings of each area and for the program overall by having multiple team members rate each program independently, meet to discuss any discrepancies in ratings, and agree on a consensus rating for each dimension. The team also created summary ratings of each area and for the program overall. To be rated fully implemented overall, a program had to receive a rating of 4 or 5 on most dimensions rated. This process could be followed by program staff who wanted to develop summary ratings. For the Baby FACES study, each program director's area rating was an average of the individual area components. Construction of the overall program rating was based on an average of each area, with the child development area weighted twice as much as the others (as in the EHSREP).

Interpretability: The results obtained from the ratings of program implementation are readily interpretable by programs serving families with infants and toddlers to show areas of programmatic strengths and weaknesses. Because the scales are tied to key dimensions of the performance standards, program management

and staff can see ways to focus program improvement efforts.

Training Support: The Early Head Start evaluation's final implementation report, *Pathways to Quality* (ACF 2002) describes the use of the checklist and rating scales in the national Early Head Start Research and Evaluation project. Copies of the rating are available in the report.

Adaptations/Special Instructions for Individuals with Disabilities: Not applicable.

Report Preparation Support: Not applicable.

References:

Administration for Children and Families.
Pathways to Quality and Full Implementation in Early Head Start Programs. Washington, DC: U.S. Department of Health and Human Services, 2003.

PROGRAM REVIEW INSTRUMENT FOR SYSTEMS MONITORING (PRISM), 2002

Authors:

American Institute for Research

Publisher:

Head Start Bureau
(866) 763-6481
puborder@headstartinfo.org

Initial Material Cost:

No costs

Representativeness of Norming Sample:

Not applicable

Languages:

English

Type of Assessment:

Comprehensive qualitative and quantitative assessment of program-level activities

Age Range and Administration Interval:

For programs serving families with children birth to age 5.

Personnel, Training, Administration, and Scoring Requirements:

Administered by Head Start Bureau personnel for program monitoring; can be used by program staff for self-review

Summary:

Initial Material Cost: 1 (< \$100)

Reliability: 1 (none described)

Validity: 1 (none described)

Norming Sample Characteristics: 1 (none described)

Ease of Administration and Scoring: 1 (none described)

Description: The Program Review Instrument for Systems Monitoring (PRISM) is both the instrument and the process used by the Administration on Children, Youth and Families to monitor Head Start programs to ensure compliance with program performance standards and other applicable regulations. The PRISM instruments are based on 17 Core Questions—9 that focus on program services and 8 that focus on management systems. The PRISM review team uses a variety of methods to learn about a program. The team members interview staff, parents, community partners, and Policy Council and governing body members, individually and in groups. They observe classrooms and family child care settings, and conduct home visits. They also complete fiscal, health and safety, and bus ride checklists. PRISM review decisions are done through consensus about the program quality.

Uses of Information: The assessment is intended to help Head Start programs identify areas that need improvements in order to continue to provide high-quality comprehensive services that meet Head Start performance standards and other regulations. The exact remedies are left to the program.

Reliability: Not described.

Validity: Not described.

Method of Scoring: Throughout the review visit, the PRISM review team, guided by the federal team leader, holds formal and informal briefings with grantee staff to report on information team members witnessed, heard, and read. During these briefings, the grantee staff members are able to provide input on the findings. At the end of the review, the review team holds an exit meeting and summarizes its findings

in three areas—Child Development and Health Services, Family and Community Partnerships, and Program Design and Management. The review team will meet to share and analyze information collected during the visit. The team will work toward building consensus on issues related to the Core Questions. The review team members will then draft a three-part report—one for each area—that summarizes the program’s strengths and areas of concerns, reviews decisions by Core Questions, and lists findings requiring corrective action. The team leader, after consulting with other Regional Office staff, will use the draft report to prepare the Official Report.

Interpretability: No information.

Training Support: The Head Start publication, *Partnership for Quality: A Grantee Guide to PRISM 2002* discusses what PRISM is

and the PRISM review process, and provides suggestions on what programs can do to prepare for the review.

Adaptations/Special Instructions for Individuals with Disabilities: Not applicable. The PRISM review criteria include evaluation of areas related to services for children with disabilities.

Report Preparation Support: Not applicable.

References:

PRISM: Program Review Instrument for Systems Monitoring of Head Start and Early Head Start Grantees—All Instruments.

PRISM: Program Review Instrument for Systems Monitoring of Head Start and Early Head Start Grantees—Partnerships for Quality: A Grantee Guide to PRISM 2002.